

Tartu Ülikool  
Loodus- ja täppisteaduste valdkond  
Matemaatika ja statistika instituut

Diana Sokurova

**Tervishoiutöötajate aruandluskoormuse vähendamine  
veereva ja kahe hetke valimidisaini abil**

Matemaatika ja statistika õppekava  
Matemaatilise statistika eriala  
Magistritöö (30 EAP)

Juhendajad: Imbi Traat, PhD  
Natalja Eigo, MSc (TAI)

Tartu 2020

# **Tervishoiutöötajate aruandluskoormuse vähendamine veereva ja kahe hetke valimidisaini abil**

Magistritöö

Diana Sokurova

**Lühikokkuvõte.** Käesoleva magistritöö eesmärk on uurida, kas iga-aastast aruandlust saab asendada uuringutega, kus küsitletakse ainult osa tervishoiuteenuse osutajaid, kuid hinnangud üldkogumi kohta soovitakse anda iga-aastaselt. Töö raames vaadeldakse kaht valikudisaini: veerev ja kahe hetke disain. Mõlemal juhul saadakse kogu üldkogum uurides vaadeldavat ja sellele eelnevat aastat. Veerevat disaini rakendatakse kahel erineval viisil – kogu üldkogumil või ainult osal üldkogumist. Viimasel juhul teatud osa üldkogumist (haiglad) täidab aruandeid iga-aastaselt. Kahe hetke disainis on olemas kattuv osa, samuti haiglad, mida vaadeldakse igal aastal, kuid hinnangud kasutavad kattuva osa põhjal leitud prognoosikonstanti. Käesolevas töös vaadeldakse kindlat aruannet – „Tervishoiutöötajad“ aastatel 2013 – 2018 ning uuritakse kolme parameetri hindamist: tervishoiutöötajate arv, tegelikult täidetud ametikohtade arv ja keskmine vanus. Tervishoiuteenuse osutajate kahe järjestikuse aasta vahel jagamiseks kasutatakse algoritmi, mis tagab võimalikult sarnased pooled teenuse liikide ja isikute arvu (või andmetabeli ridade) suhtes. Lahendatakse ka küsimus, kuidas võtta arvesse üldkogumi iga-aastast muutumist uute ja töö lõpetanud asutuste suhtes. Parimateks hinnanguteks osutusid kattuva osaga veereva disaini hinnangud, kus haiglad esitasid aruandeid iga-aastaselt.

**CERCS teaduseriala:** P160 Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika

**Märksõnad:** valikuuringud, valikuteooria, hindamine, täpsus, veerev disain, kahe hetke disain

## **Reducing the Reporting Burden on Health Professionals by Rolling and Two Occasions Designs**

Master's thesis

Diana Sokurova

**Abstract.** The aim of this master's thesis is to investigate whether annual statistical reporting can be replaced by surveys in which only a part of healthcare providers is interviewed, but it still provides estimates for the whole population on an annual basis. Within the framework of

the thesis, two types of designs are considered: rolling and two occasions designs. In both cases, the whole population is obtained by surveying the observed and the preceding years. Rolling design is applied in two different ways, i.e. to the whole population or only to a part of the population. In the latter case, a certain part of the population (hospitals) provides data annually. Similarly, hospitals are an overlapping part in the two occasions design, observed each year. Differently from the rolling designs, the estimates here use a prediction constant found on the overlapping part. In this thesis, we study a particular statistical report – “Healthcare Professionals” for years 2013 – 2018 and examine the estimations of three parameters: the number of healthcare professionals, the number of positions actually filled and the average age. An algorithm is used to divide healthcare providers into two consecutive years, ensuring the most similar groups in terms of service types and number of persons (or rows in the data table). The question of how to take into account the annual change in the healthcare providers population for new and closed facilities is also resolved. The best estimates were received by overlapping rolling designs, where hospitals provided data annually.

**CERCS research specialisation:** P160 Statistics, operation research, programming, actuarial mathematics

**Keywords:** survey sampling, sample survey theory, estimation, accuracy, rolling samples, sampling on two occasions

# Sisukord

Sissejuhatus .....	5
1 Üldkogum ja huvialused parameetrid.....	7
1.1 Tervishoiutöötajate aruandes esitatav teave .....	7
1.2 Asutuste informatsioon .....	8
1.3 Avaldatava informatsiooni kirjeldus .....	8
1.4 Parameetrite arvutamine.....	9
2 Metoodika aruandluskoormuse vähendamiseks .....	13
2.1 Veereva disaini teoreetiline taust .....	13
2.2 Veereva disaini rakendus antud töös.....	14
2.2.1 Veerev disain 1 .....	14
2.2.2 Veerev disain 2.....	15
2.3 Kahe hetke disaini teoreetiline taust.....	16
2.4 Kahe hetke disaini rakendus antud töös .....	17
2.4.1 Kahe hetke disain 1 .....	18
2.4.2 Kahe hetke disain 2 .....	19
3 Ettevalmistus praktiliseks ülesandeks .....	21
3.1 Kihtide moodustamine .....	21
3.2 Algoritm üldkogumi jagamiseks kaheks osaks .....	22
3.3 Aastatega muutuv andmestik .....	23
3.4 Moodustatud valimite ülevaade .....	24
4 Saadud tulemuste analüüs .....	27
4.1 Tervishoiutöötajate arvu hinnangud.....	27
4.2 Tegelikult täidetud ametikohtade hinnangud .....	31
4.3 Hinnangud keskmisele vanusele .....	35
Kokkuvõte .....	37
Kasutatud kirjandus.....	39
Lisa 1. R kood asutuste pooleks jagamiseks .....	40
Lisa 2. Hinnangud erinevate meetodite ja tunnuste korral .....	42

## Sissejuhatus

Tervise andmete kogumine ja riikliku tervisestatistika avaldamine allub reale määrustele ja seadustele, mida on kokkuvõtlikult kirjeldatud varem avaldatud Tervise Arengu Instituudi (TAI) analüüsis [1].

Igal Terviseametis tegevusloa saanud tervishoiuteenuse osutajal (edaspidi TTO, tervishoiuasutus või asutus) on tervishoiuteenuste korraldamise seaduse § 47 alusel kohustus esitada tervishoiustatistilisi aruandeid valdkonna eest vastutava ministri poolt määratud asutusele, mis on TAI [2].

Tervishoiustatistilisi aruandeid täidavad TTOde raamatupidajad, personali spetsialistid, haigla statistikud või väikeste asutuste korral ka arstid ja õed. Iga-aastaste aruannete arv on suur, viimastel aastatel kuni 13. TTOde andmete põhjal saadud statistilised näitajad avaldatakse Tervisestatistika ja terviseuuringute andmebaasis [3]. Nende abil seiratakse Eesti tervishoiusüsteemi, langetatakse tervishoiupoliitilisi otsuseid, kirjutatakse teadusartikleid, tehakse analüüse ning neid edas-tatakse rahvusvahelistele organisatsioonidele, nt Euroopa Komisjoni statistikaamet (*Eurostat*) ja Maailma Terviseorganisatsioon (*WHO*). Andmeid toodetakse iga aasta kohta ning oluline on nende võrreldavus.

Samas on statistika kogumisel üheks ülesandeks hoida TTOde aruandluskoormus võimalikult madalal. TAI uurib aktiivselt võimalusi vähendamaks tervishoiuasutuste ja TAI töötajate koormust. Koormuse vähendamisel tuleb siiski tagada andmete olemasolu ja kvaliteet. Käesolevas magistritöös uuritakse kahte meetodit aruandluse esitamise sageduse vähendamiseks TTO jaoks, säilitades seejuures vajaliku aastase statistika. Nendeks meetoditeks on

- 1) hindamine veereva disaini (i.k. *rolling sampling*) korral, kus mahukas loendus asendatakse väiksemate lõikumate uuringutega [4];
- 2) hindamine kahe hetkega disaini (i.k. *sampling on two occasions*) korral, kus lubatakse valimite lõikumist [5].

Antud magistritöö eesmärk on uurida, kas iga-aastast aruandlust saab asendada eespool kirjeldatud meetodite või nende modifikatsioonidega. Eesmärk on välja selgitada, milline meetod sobib selleks paremini.

Käesolevas töös vaadeldakse kindlat aruannet – „Tervishoiutöötajad“, mida peavad esitama kõik tervishoiuasutused kõigi nendes töötavate tervishoiutöötajate kohta.

Kuna tervishoiutöötajad võivad töötada mitmel ametil ja mitmes asutuses, siis meetodite rakendamise muudab keerukaks andmestiku struktuur. Nimelt võib sama tervishoiutöötaja esineda andmestikus mitmekordselt, mida tuleb parameetrite arvutamisel arvesse võtta.

Magistritöös väljapakutud meetodite käitumist uuritakse reaalsel andmetel. Selleks kasutatakse aruandega “Tervishoiutöötajad” kogutud andmeid aastatest 2013–2018. Need andmed sisaldavad informatsiooni TTOde kohta. Samuti millistes ametites, millise vanuse, soo ja töökoormusega tervishoiutöötajad töötavad nendes asutustes.

Käesolev magistritöö koosneb neljast peatükist. Esimeses peatükis kirjeldatakse aruandes „Tervishoiutöötajad“ kogutavaid ja avaldatavaid andmeid. Valitud metoodikat aruandliskoormuse vähendamiseks tutvustatakse teises peatükis. Kolmandas peatükis esitatakse praktilised sammud TTOde jagamiseks eri aastate vahel. Vaadatakse kahte erinevat jagamisviisi. Selgitatakse, kuidas võtta arvesse üldkogumi iga-aastast muutumist uute ja töö lõpetanud asutuste suhtes. Valitud parameetrite hinnangud ja lähendusvead tuuakse koos illustratsioonide ja kommentaaridega neljandas peatükis. Valikudisainide rakendamiseks kasutatakse tarkvara R ja hinnangute leidmiseks tarkvara STATA.

# 1 Üldkogum ja huvialused parameetrid

Tegemist on kõikse uuringuga, mille iga-aastane üldkogum koosneb “Tervishoiutöötajate” aruannetest ja kõikidest tervishoiuteenust pakutavatest asutustest. Allpool kirjeldatakse, mis on üldkogumi objektiks ja tunnusteks ning kuidas arvutatakse parameetrid.

## 1.1 Tervishoiutöötajate aruandes esitatav teave

“Tervishoiutöötajate” aruanne on üks aruannetest, mida iga tervishoiuteenuse osutaja on kohustatud TAIle esitama. Sellega kogutakse tervishoiutöötajate andmeid novembrikuu kohta. Antud aruandes tuuakse välja iga tervishoiutöötaja kohta tema amet antud TTOs. Kui isik töötab asutuses mitme erineva lepingu alusel erineval ametikohal või lisaks teises asutuses, siis näidatakse selle isiku andmeid tema kõikide ametikohtade kohta erinevatel ridadel ehk mitmekordselt. TAIs koostatakse andmestik, kuhu on koondatud kõikide TTOde aruanded.

Andmestikus on unikaalseks objektiks kolmik: asutus, isik ja isiku amet. See kolmik esineb andmestikus ilma kordusteta. Andmestiku iga rida on selle kolmiku mõttes erinev objekt. Objektide arv üldkogumis on seega andmestiku ridade arv. Andmestiku tunnused on välja toodud tabelites 1 ja 2.

Tabel 1. Aruandega “Tervishoiutöötajad” kogutav teave

Tunnuse nimi	Selgitus	Võimalikud väärtused või näited
1. aasta	Aruandeaasta	2013–2018
2. id_kood	TTO unikaalne kood	TAIs registreeritud asutuse järjekorra number
3. nimi	TTO nimi	Näide: SA Tartu Ülikooli Kliinikum
4. isiku_id_num	Eesti elanike unikaalne isikukood	Üheteistkümnest numbrist koosnev kood
5. ametikood	Töötaja ametikood. Ametid kodeeritakse vastavalt Statistikaameti ja TAI poolt väljatöötatud ametite klassifikaatorile kaheksakohalise koodiga [6].	Näited: 22110201 Perearst 22122702 Psühhiaater 22210502 Õde (üldõde) 22610102 Hambaarst Kokku on 72 tervishoiutöötajate ametit.
6. juht	Kas on juht? Juhiks loetakse isikut, kellel on alluvaid, kes tegeleb protsessi juhtimisega ja vastutab üksuse/osakonna/asutuse töö eest.	on juht/omanik, juht, omanik, FIE või ei ole
7. resident	Kas on residentuuri lepinguga?	0–Ei ole; 1–Jah.
8. koormus_leping	Töölepingus määratud koormus	Min 0; maks 2; keskmine 0,799.
9. tegelik_koormus	Tegelikult novembris töötatud tundide arvust tulenev koormus	Min 0; maks 2,88; keskmine 0,725.
10. sh_ülekoormus	Erinevus tegeliku ja lepingulise koormuste vahel	Min 0; maks 1,488; keskmine 0,022.

## 1.2 Asutuste informatsioon

Lisaks kogub TAI pidevalt informatsiooni tegutsevate TTOde kohta ja koostab iga-aastase TTOde statistilise registri. Tööandmestiku rikastamiseks lisatakse sellest registrist aruande andmete juurde järgmised tunnused (tabel 2) .

Tabel 2. Informatsioon asutuste kohta

Tunnuse nimi	Selgitus	Võimalikud väärtused
<b>1. regkood</b>	TTO äriregistrikood	Kaheksakohaline unikaalne kood, mis antakse asutusele registreerimisel
<b>2. maakond suurlinn</b>	TTO tegevuskoht. Kodeerimine vastab Eesti haldus- ja asustusjaotuse klassifikaatorile [7].	Tegevuskohad on toodud Tabelis 5.
<b>3. teenus</b>	TTO osutatav peamine tervishoiuteenus. Kodeerimine vastab TAI poolt tehtud tervishoiuteenuse osutajate liigitusele [8].	10 – Haigla 20 – Perearstiabiasutus 25 – Eriarstiabiasutus 30 – Hambaraviasutus 40 – Kiirabiasutus 50 – Taastusraviasutus 51 – Diagnostikaasutus 53 – Õendusabiasutus 90 – Muu asutus 99 – Üldarst, v.a. perearst
<b>4. omvorm</b>	Omandivorm	riik, kohalik omavalitsus, eesti eraõiguslik isik, välismaa eraõiguslik isik, muu
<b>5. haiglaliik</b>	Haiglaliik	0 – Ei ole haigla 11 – Piirkondlik 12 – Keskhaigla 13 – Üldhaigla 14 – Taastusravihaigla 15 – Õendushaigla 16 – Erihaigla 17 – Kohalik haigla
<b>6. haiglavõrk</b>	Kas on haiglavõrgu arengukava haigla?	ei ole haiglavõrgu arengukava haigla, on haiglavõrgu arengukava haigla
<b>7. HP</b>	Tervishoiuteenuse pakkujate liigitus	Vastab tervishoiuteenuste pakkujate koondklassifikatsioonile (SHA 2011, ICHA-HP) [9].
<b>8. omvormbi</b>	Omandivorm (avalik/era)	avalik sektor, erasektor

## 1.3 Avaldatava informatsiooni kirjeldus

Iga-aastaste aruannete koondtulemused avaldatakse tervisestatistika ja terviseuuringute andmebaasis [3]. Avaldatavateks näitajateks (edaspidi parameetriteks) on tervishoiutöötajate arv, keskmine vanus, täidetud ametikohad lepingulise koormuse järgi, tegelikult täidetud ametikohad, ületundidest tulenev täiendav täidetud ametikohtade arv erinevate tunnuste lõikes.



Järgnevalt on välja toodud kõigi avaldatavate parameetrite loetelu<sup>1</sup>, sulgudes on tabeli kood andmebaasis.

1. Töötavate isikute absoluutarv ja kordaja 100 000 elaniku kohta (sümboliga \* on märgitud need tabelid, kus väljastatakse ainult absoluutarv):
  - a) ameti lõikes (THT001);
  - b) ameti, soo, vanuserühma lõikes\* (THT002);
  - c) ametiala ja TTO liigi lõikes (THT006);
  - d) ametiala ja TTO omaniku liigi (avalik või era sektor) lõikes (THT007);
  - e) ametiala ja lepingulise koormuse lõikes\* (THT008) ;
  - f) perearstiabiastutuses ametiala ja maakonna järgi (THT009);
  - g) hambaraviteenuse osutamisega seotud ametiala ja maakonna järgi (THT010);
  - h) TTO ja omaniku liigi järgi\* (THT013).
2. Täidetud ametikohad lepingulise koormuse järgi, tegelikult täidetud ametikohad, ületundidest tulenev täiendav täidetud ametikohtade absoluutarv ja kordaja 100 000 elaniku kohta:
  - a) ameti lõikes (THT001);
  - b) ametiala ja TTO liigi lõikes (THT006);
  - c) ametiala ja TTO omaniku liigi (avalik või era sektor) lõikes (THT007);
  - d) perearstiabiastutuses ametiala ja maakonna järgi (THT009);
  - e) hambaraviteenuse osutamisega seotud ametiala ja maakonna järgi (THT010).
3. Töötavatest arst- ja hambaarst-residentidest absoluutarv ameti järgi (THT003).
4. Tervishoiutöötajate keskmine vanus:
  - a) soo ja ameti lõikes (THT004);
  - b) ametiala ja TTO liigi lõikes (THT005).

#### **1.4 Parameetrite arvutamine**

Parameetrite arvutamise muudab keerukaks andmestiku struktuur, kus sama isik võib esineda korduvalt, st andmestiku mitmel real.

Maailmas on teisigi andmebaase, mis sisaldavad ühte isikut mitu korda. Näiteks on Tööregistris inimesed mitme erineva töökohaga, vähiregistris erineva vähitüübiga, esmahaigestumise statistikas erineva diagnoosiga. Kui on soov hinnata inimeste arvu osakogumite lõikes, näiteks

---

<sup>1</sup> Aruandest „Tervishoiutöötajad“ avaldatav teabe on andmebaasis „tervishoiu ressursi ja nende kasutamine“ ja „Tervishoiutöötajad“ alamkaustates  
[https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas\\_04THressursid\\_05Tootajad/?tablelist=true](https://statistika.tai.ee/pxweb/et/Andmebaas/Andmebaas_04THressursid_05Tootajad/?tablelist=true)

inimeste arvu erinevate vähipaikmete lõikes, siis sama isik saab olla mitmes osakogumis, kui tal on diagnoositud mitu vähki. Isikute arvu leidmiseks osakogumite lõikes on erinevaid võimalusi. Järgnevalt tuuakse välja kolm võimalust:

- 1) iga osakogumi jaoks koostatakse oma andmestik, kust eemaldatakse kordused ja seejärel arvutatakse inimeste arv;
- 2) andmestik viiakse laia formaati nii, et ridades on inimesed (unikaalselt esitatud) ja veergudes indikaatortunnused iga osakogumi jaoks;
- 3) lisatakse tunnus, mis järjestab isikute kordused antud osakogumis ja seejärel arvutatakse inimeste arv, kelle järjekorra number on 1 antud grupis.

Kõigi nende meetoditega kaasneb probleem, et inimeste arvud osakogumites ei summeeru inimeste arvuks üldkogumis. Näiteks eri tüüpi vähki põdevate inimeste arvud ei anna kokku vähihaigete üldarvu.

Tervise Arengu Instituudis kasutakse tervishoiutöötajate aruannete töötlemisel kolmandat varianti. Suure osakogumite arvu ja mahukate andmefailide tõttu on see variant neist kolmest kõige efektiivsem.

Järgnevalt paneme valemite abil kirja selle, kuidas TAI ülal loetletud parameetreid arvutab. Olgu üldkogumiks ühe aasta andmestik tähisega  $U = \{1, 2, \dots, i, \dots, N\}$ , kus objektiks  $i$  on andmestiku rea number ja  $N$  on ridade arv. Olgu  $M$  isikute arv. Isikud üldkogumis  $U$  võivad olla kordustega. Olgu osakogum  $U_d$  nende objektide (ridade) hulk, mis rahuldavad teatud tingimust  $d$ . Näiteks  $U_{naised}$  sisaldab naistöötajaid. Sama isik võib ka siin korduvalt esineda, kui ta esineb mitu korda andmestikus.

Parameetrid, mis on seotud kogukoormusega (täidetud ametikohad lepingulise koormuse järgi, tegelikult täidetud ametikohad, ületundidest tulenev täiendav täidetud ametikohtade arv) põhinevad üldkogumi  $U$  lihtsatel kogusummadel, hoolimata isikute kordustest:

$$t_y = \sum_{i \in U} y_i = \sum_{i=1}^N y_i, \quad (1.1)$$

kus  $y_i$  on tunnuse  $y$  väärtus üldkogumi objektil  $i$ . Samade parameetrite leidmisel osakogumis  $U_d$  kasutatakse samuti lihtsaid kogusummasid, seekord üle osakogumi,

$$t_{yd} = \sum_{i \in U_d} y_i. \quad (1.2)$$

Valemi (1.1) abil leitakse näiteks summaarne lepingujärgne koormus, kus  $y_i$  on tunnuse *koormus leping* väärtus objektil  $i$ . Summaarne tegelik koormus naistel leitakse aga valemi (1.2) abil,  $\sum_{i \in U_{naised}} \text{tegelik koormus}_i$ .

Tervishoiutöötajate arvu arvutamisel ja keskmise vanuse leidmisel on oluline, et isikud oleksid unikaalsed. Selleks moodustatakse osakogumid  $U_{unik}$  ja  $U_{d-unik}$ , kus  $d$  on huvipakkuv rühm, mille lõikes vaadeldakse üksnes unikaalseid isikuid. Üle  $U_{unik}$  summeerides leitakse tervishoiutöötajate koguarv, mida saab lisaks leida ka soo ja vanuse lõikes, sest need on isikut kirjeldavad tunnused. Samas saab luua unikaalsuse indikaatortunnuse  $I_{unik}(i)$ , mis on 1, kui isik reas  $i$  esineb üldkogumis esimest korda ja 0, kui isik esines üldkogumis juba varem. Siinjuures eeldame, et üldkogum vaadatakse läbi rea kaupa andmestiku esimesest reast viimaseni. Osakogumi jaoks defineeritakse grupi  $d$  indikaatortunnus  $I_{d-unik}(i)$  järgmiselt:

$$I_{d-unik}(i) = \begin{cases} 1, & \text{kui } d(i) = 1 \text{ ja } \nexists j < i: d(j) = 1 \text{ ja } isik(j) = isik(i) \\ 0, & \text{kui } d(i) = 0 \text{ või } \exists j < i: d(j) = 1 \text{ ja } isik(j) = isik(i) \end{cases} \quad (1.3)$$

kus  $d(i) = 1$ , kui tingimus  $d$  kehtib real  $i$ , ja  $d(i) = 0$ , kui ei kehti ning  $isik(i)$  on real  $i$  oleva isiku isikukood  $i, j \in U$ . Nüüd on võimalik arvutada inimeste koguarvu ja tunnuse *vanus* kogusummat nii üldkogumis kui ka osakogumites, kasutades järgmisi valemeid:

$$\begin{aligned} t_{y,unik} &= \sum_{i \in U_{unik}} y_i = \sum_{i=1}^N I_{unik}(i) \cdot y_i, \\ t_{y,d-unik} &= \sum_{i \in U_{d-unik}} y_i = \sum_{i=1}^N I_{d-unik}(i) \cdot y_i, \end{aligned} \quad (1.4)$$

kus  $y_i = 1$ , kui arvutatakse inimeste koguarvu ja  $y_i = \text{vanus}_i$ , kui tunnuse *vanus* kogusummat.

Tervishoiutöötajate arv  $M$  arvutatakse järgmiselt:

$$M = t_{1,unik} = \sum_{i=1}^N I_{unik}(i) \cdot 1.$$

Osakogumis  $U_d$  arvutatakse tervishoiutöötajate arv  $M_d$  valemiga:

$$M_d = t_{1,d-unik} = \sum_{i=1}^N I_{d-unik}(i) \cdot 1. \quad (1.5)$$

Keskmine vanus arvutatakse järgmiselt:

$$keskmine\ vanus = \frac{t_{vanus,unik}}{t_{1,unik}} = \frac{\sum_{i=1}^N I_{unik}(i) \cdot vanus_i}{\sum_{i=1}^N I_{unik}(i)}.$$

Keskmise vanuse leidmisel osakogumis  $d$  kasutatakse selles valemis indikaatorit  $I_{d-unik}(i)$ .

Järgnevalt vaatame arvutusi tabelis 3 toodud näite abil (asutuse ja isiku ID on siin vastavalt 1- ja 3-kohaline number). Isik 001 töötab hambaarstina asutuses 1 ja 2 ning erakorralise meditsiini õena asutuses 3, aga isik 002 töötab asutuses 4 kolmel erineval ametikohal. Antud näites on kokku 2,35 täidetud ametikohta lepingulise koormuse järgi, hambaarstidel residentidega on 1 ja õdedel on 1,35. Isikuid on aga 2, kellest 1 on hambaarst ja mõlemad on õed. Keskmine vanus antud näites on

$$\frac{\sum_{i=1}^6 I_{unik}(i) \cdot vanus_i}{\sum_{i=1}^6 I_{unik}(i)} = \frac{43 \cdot 1 + 35 \cdot 1}{1 + 1} = \frac{78}{2} = 39.$$

Õdede keskmine vanus on sama, 39, aga hambaarstide keskmine vanus on 43.

Tabel 3. Näited mitmel ametikohal töötavatest isikutest, andmed 2018.a

Rea nr	Asutuse ID	Isiku ID	Vanus	Koormus leping	Amet	Ametigrupp	$I_{unik}$	$I_{hambaarsid-unik}$	$I_{õed-unik}$
1	1	001	43	0,5	Hambaarst	Hambaarstid residentidega	1	1	0
2	2	001	43	0,5	Hambaarst	Hambaarstid residentidega	0	0	0
3	3	001	43	0,5	Erakorralise meditsiini õde	Õed	0	0	1
4	4	002	35	0,25	Õde (üldõde)	Õed	1	0	1
5	4	002	35	0,5	Operatsiooni õde	Õed	0	0	0
6	4	002	35	0,1	Koduõde	Õed	0	0	0
<b>Kokku</b>				<b>2,35</b>	<b>Kokku</b>		<b>2</b>	<b>1</b>	<b>2</b>

## 2 Metoodika aruandluskoormuse vähendamiseks

Kirjeldame kahte meetodit – veerevat disaini ja kahe hetke disaini. Vaatleme nende rakendust antud töös koos modifikatsioonidega. Koos modifikatsioonidega on vaatluse all 4 meetodit.

Aruandluses on huvipakkuvateks parameetriteks enamasti kogusummad. Seega keskendutakse töös kogusumma hindamisele. Vaatleme erinevaid käsitlusi vajavaid kogusummasid, mis nõuavad või ei nõua unikaalsete isikute väljaselgitamist:

- 1) isikute arv, mis ei tohi sisaldada korduseid;
- 2) täidetud ametikohtade arv (lepinguline, tegelik ja ülekoormus);
- 3) vanuse kogusumma, mida kasutatakse keskmise vanuse leidmiseks ja mis ei tohi sisalda isikute korduseid.

### 2.1 Veereva disaini teoreetiline taust

Veerev valikudisain (i.k. *rolling sampling*) tähendab  $k$  erineva (mittekattuva) perioodilise valimi moodustamist, millest igaüks on tõenäosuslik valim üldkogumist osakaaluga  $f = \frac{1}{F}$  ( $F$  on perioodide arv), nii et  $k$  perioodi kumuleerimisel saadakse kogu üldkogumi valim osakaaluga  $f = \frac{k}{F}$ . Kui  $k = F$ , saadakse kogu üldkogum [4].

Veerevat disaini on rakendatud loenduste korral, kus ka loendusaastate vahepeal on vaja rahvastikunäitajaid hinnata [4]. Näiteks olgu loendustel 10-aastane tsükel. Eeldame, et käesolevale aastale eelneva aasta  $i$  ( $i = 0; 1; \dots; 9$ ) keskmine on  $\bar{y}_i$ . Käesoleva aasta üldkeskmine on  $\bar{y} = \sum_{i=0}^9 W_i \bar{y}_i$ , kus  $W_i$  tähistab kaalu, mis on määratud kümnele aastale ja  $\sum_{i=0}^9 W_i = 1$ . Järgnevalt on võimalik koostada neli mudelit käesoleva aasta keskmise arvutamiseks [10]. Olgu vastavad keskmised:  $\bar{y}_c$ ,  $\bar{y}_f$ ,  $\bar{y}_e$  ja  $\bar{y}_w$ .

1. Loendusel (*census*) põhinev arvutamine:  $\bar{y}_c = \bar{y}_0$ , kus  $W_0 = 1$  ja ülejäänute  $i > 0$  korral  $W_i = 0$ . See sarnaneb praktikaga, kus kõik hinnangud baseeruvad samale aastale, millal toimus loendus.
2. Viimasel (*final*) aastal põhinev arvutamine:  $\bar{y}_f = \bar{y}_9$ , kus  $W_9 = 1$  ja ülejäänud  $i < 9$  korral  $W_i = 0$ . Seda saab kasutada üleriigiliste ja suurte piirkondade hinnangute puhul, kus aeg mõjutab hinnangute täpsust ja valimi maht on piisavalt suur. Seda saab kasutada näiteks epideemiate, aktsiahindade, tööpuuduse jm hindamisel, kus ebaregulaarsed kõikumised on iseloomulikumat kui stabiilsed muutused.

3. Võrdsete kaaludega (*e-equal*) arvutamine:  $\bar{y}_e = \sum W_i \bar{y}_i$ , kus  $W_i = 0,1$  iga  $i$  korral. See sobib (suhteliselt) muutumatutes stabiilsetes olukordades. Sageli väidetakse, et sellist stabiilset mudelit on parem kasutada vananenud kümneaastaste loenduste asemel.
4. Monotoonselt mittekahenevate kaaludega  $W_i$  arvutamine:  $\bar{y}_w = \sum W_i \bar{y}_i$  (*w-weight*), kus  $W_0 \leq W_1 \leq \dots \leq W_8 \leq W_9$ . Kasvukõverat võib kindlaks määrata mõni mudel, empiirilised andmed või nende kombinatsioon.

Keskmised  $\bar{y}_f$  ja  $\bar{y}_e$  on triviaalsed, aga nad on paremad kui  $\bar{y}_c$ , sest kasutavad värskemaid andmeid. Samal ajal mingi monotonne kasvamine nagu  $W_{i+1} = W_i + k$  ( $k > 0$ ) või  $W_{i+1} = cW_i$  ( $c > 1$ ) võib olla parem kui  $W_i = 0.1$ .

## 2.2 Veereva disaini rakendus antud töös

Käesolevas magistritöös jagatakse üldkogum kaheks osaks, millest üks osa täidab aruandeid ühel ja teine teisel aastal. Seega perioodide arv  $F = 2$ . Modifikatsioonina vaadeldakse võimalust, kus teatud osa üldkogumist täidab aruandeid igal aastal. Üldkogumi jagamine ei ole juhuslik, vaid toimub kindla algoritmi alusel, mida kirjeldatakse punktis 3.2. Eesmärk on saada kihtidesse võimalikult võrdsed töötajate arvud mõlemal aastal. Kihtideks on valitud TTO liigid (haigla, perearstiabi, hambaravi jne).

Arvestada tuleb siinjuures sellega, et üldkogum muutub igal aastal. Mõned tervishoiuteenuse osutajad lõpetavad tegevuse ja teised alustavad. Muutuse arvesse võtmist kirjeldatakse punktis 3.3.

### 2.2.1 Veerev disain 1

Selleks, et saada igal aastal hinnangud üldkogumile, vaadatakse kahte aastat: vaadeldav ja eelnev. Mõlemal aastal on vaid osade objektide andmed, mis pannakse kokku üheks üldkogumiks, mis koosneb kahest mittelõikuvast poolest. Olgu vaadeldava aasta pool  $U_1$  ja sellele eelneva aasta pool  $U_0$ . Tervishoiutöötajate arvu leidmiseks kasutame kahe kogusumma lineaarkombinatsiooni.

$$\hat{M} = 2(w_0 t_{1,d_0-unik} + w_1 t_{1,d_1-unik}), \quad (2.1)$$

kus  $w_0$  on eelneva aasta kaal ja  $w_1$  vaadeldava aasta oma,  $w_0 \leq w_1$  ja  $w_0 + w_1 = 1$ . Kordaja 2 on proportsionaalsuse kordaja, sest sulgudes olev osa hindab kogusumma vaid poole üldkogumi jaoks. Esimese ja teise poole töötajate arvude kogusummad  $t_{1,d_0-unik}$  ja  $t_{1,d_1-unik}$  arvutatakse valemi (1.4) abil. Siin  $d_0$  ja  $d_1$  on osakogumeid defineerivad tingimused, kus  $d_1$  tähendab, et

antud isik esineb pooles  $U_1$  ja  $d_0$  tähendab, et isik esineb pooles  $U_0$ , aga ei esine pooles  $U_1$ . Esimene indeks 1 kogusummades viitab sellele, et isikute arvu hindamisel on uuritav tunnus  $y = 1$ .

Lisaks on meil vaadeldaval aastal küsitletud pooldest teada selle tegelik isikute arv. Tähistame selle  $M_1$ . Siit saame hinnata, mitu isikut ei sattunud küsitletud poolde  $\hat{M}_0 = \hat{M} - M_1$ . Nüüd saame kasutada isikute arvu kaaludena teiste kogusummade arvutamisel:

$$\hat{t}_y = \frac{\hat{M}}{\hat{M}_0} w_0 t_{y_0} + \frac{\hat{M}}{M_1} w_1 t_{y_1}, \quad (2.2)$$

kus  $t_{y_0}$  on tunnuse  $y$  kogusumma eelmises pooles ja  $t_{y_1}$  on tunnuse  $y$  kogusumma vaadeldavas pooles (valem (1.2)).

Vanuse kogusumma leitakse üle unikaalsete isikute. Seega selle hinnang avaldatakse valemiga

$$\hat{t}_{vanus} = \frac{\hat{M}}{\hat{M}_0} w_0 t_{vanus_0-unik} + \frac{\hat{M}}{M_1} w_1 t_{vanus_1-unik}, \quad (2.3)$$

kus  $t_{vanus_0-unik}$  ja  $t_{vanus_1-unik}$  on unikaalsete isikute vanuse kogusumma vastavalt eelmise ja vaadeldava aasta pooles, mis arvutatakse valemiga (1.4).

### 2.2.2 Veerev disain 2

Eelmises peatükis rakendati veerevat disaini kogu üldkogumile. Järgnevalt vaatleme osalist veerevat disaini, kus haiglad täidavad aruandeid iga-aastaselt, aga veerevat disaini rakendatakse ülejäänud TTOdele ehk mittehaiglatele.

Haiglates töötavate unikaalsete tervishoiutöötajate arv  $M_H$  leitakse osakogumi valemist (1.5), kus osakogumit defineeriv tunnus  $d$  näitab, kas isik töötab haiglas. Haiglas mittetöötavate isikute arvu  $M_{MH}$  leidmiseks lisatakse valemisse (2.1) tingimus, et antud isik ei tööta haiglas,

$$\hat{M}_{MH} = 2(w_0 t_{1,MH-d_0-unik} + w_1 t_{1,MH-d_1-unik}), \quad (2.4)$$

kus  $MH - d_1$  tähendab nende isikute osakogumit, kes esinevad pooles  $U_1$  ja ei tööta haiglas ning  $d_0$  tähendab isikute osakogumit, kes esinevad pooles  $U_0$ , aga ei esine pooles  $U_1$  ega tööta haiglas.

Veereva disaini 2 korral on tervishoiutöötajate arvu hinnang,

$$\hat{M} = M_h + \hat{M}_{MH}. \quad (2.5)$$

Analoogiliselt leitakse ka koormuse kogusumma hinnang,

$$\hat{t}_y = t_{y_H} + \hat{t}_{y_{MH}}, \quad (2.6)$$

kus  $t_{y_H}$  on tunnuse  $y$  kogusumma haiglates ja  $\hat{t}_{y_{MH}}$  on tunnuse  $y$  kogusumma hinnang mittehaiglates.

Viimane arvutatakse järgneva valemiga

$$\hat{t}_{y_{MH}} = \frac{\hat{M}_{MH}}{\hat{M}_{0-MH}} w_0 t_{y_{MH-0}} + \frac{\hat{M}_{MH}}{M_{1-MH}} w_1 t_{y_{MH-1}}, \quad (2.7)$$

kus  $M_{1-MH}$  on mittehaiglates töötavate isikute arv vaadeldaval aastal ja  $\hat{M}_{0-MH} = \hat{M}_{MH} - M_{1-MH}$  on vastav hinnanguline arv eelneval aastal. Suurused  $t_{y_{MH-1}}$  ja  $t_{y_{MH-0}}$  on tunnuse  $y$  kogusummad vastavalt vaadeldava ja sellele eelneva aasta pooles.

Vanuse kogusumma leitakse valemiga

$$\hat{t}_{vanus} = t_{vanus_H} + \hat{t}_{vanus_{MH}}, \quad (2.8)$$

kus  $t_{vanus_H}$  on haiglas töötavate unikaalsete isikute vanuse kogusumma ja  $\hat{t}_{vanus_{MH}}$  on nende unikaalsete isikute vanuse kogusumma hinnang, kes ei tööta haiglas.

### 2.3 Kahe hetke disaini teoreetiline taust

Kahe hetke disain (*Sampling on Two Occasions*) on kirjeldatud raamatus Model Assisted Survey Sampling [5]. Kahe hetke disain on protsess, kus tõenäosuslikke valimeid võetakse üldkogumist kahel erineval ajahetkel. Lubatud on valimite kattuvust. Kattuvast osas on samu objekte küsitatud nii esimesel kui ka käesoleval hetkel. Eesmärk on käesoleva hetke kogusumma või keskmise hindamine. Kahe hetke valimid ei kata kogu üldkogumit. Töös kasutatud tähistused on pärit allikast [5]:

$U = \{1, \dots, i, \dots, N\}$  – üldkogum, kus  $N$  on üldkogumi maht,

$z$  – uuritav tunnus esimesel hetkel,

$y$  – uuritav tunnus teisel hetkel,

$s_a$  – esimese hetke valim,

$s_m$  – kattuv (matched) valim, mis on võetud valimist  $s_a$ ,

$s_u$  – mittekatuv (unmatched) valim, mis on võetud täiendvalimist  $s_a^c$ ,

$s = s_m \cup s_u$  – teise hetke valim.

Olgu teise hetke väärtused  $y_k$  sellised, et nad on hästi lähendatavad suurusega  $y_k^0 = Kz_k$ , kus  $K$  on teatud konstant.  $K$  leitakse varasemaid andmeid kasutades või antakse eksperdi poolt. Kat-



tuval valimil saame leida erinevuse  $D_k = y_k - y_k^0$ , mille abil saame nihketa hinnangu teise hetke kogusummale valimis  $s_a$ ,

$$\hat{t}_1 = \hat{t}_{y^0 s_a} + \hat{t}_{D s_m}, \quad (2.9)$$

mis arvutatakse suuruse  $y_k^0$  kogusumma hinnangu ja erinevuse  $D_k$  kogusumma hinnangu abil.

Need hinnangud avaldatakse järgmiselt:

$$\hat{t}_{y^0 s_a} = \sum_{s_a} \frac{y_k^0}{\pi_{ak}},$$

$$\hat{t}_{D s_m} = \sum_{s_m} \frac{D_k}{\pi_{ak} \pi_{k|s_a}},$$

kus  $\pi_{ak}$  ja  $\pi_{k|s_a}$  on objekti  $k$  kaasamistõenäosused vastavalt valimitesse  $s_a$ , ja teisel hetkel valimisse  $s_m$  tingimusel, et ollakse valimis  $s_a$ .

Teine nihketa hinnang praegusele kogusummale on  $y_k$  kogusumma hinnang teisel hetkel, mis saadakse mittekattuvast valimist,

$$\hat{t}_2 = \hat{t}_{y s_u} = \sum_{s_u} \frac{y_k}{\pi_{ak}^c \pi_{k|s_a}^c}, \quad (2.10)$$

kus  $\pi_{ak}^c$  ja  $\pi_{k|s_a}^c$  on kaasamistõenäosused vastavalt esimese hetke täiendvalimisse  $s_a^c$ , ja teisel hetkel mittekattuvasse valimisse  $s_u$  tingimusel, et ollakse esimese hetke täiendvalimis.

Kasutades nende lineaarset kombinatsiooni, saame uue nihketa hinnangu,

$$\hat{t}_y = w_1 \hat{t}_1 + w_2 \hat{t}_2,$$

kus  $w_1$  ja  $w_2$  on mittenegatiivsed kaalud, nii et  $w_1 + w_2 = 1$ . Hinnangut  $\hat{t}_y$  nimetatakse liithinnanguks, kuna ta ühendab kattuva valimi ja mittekattuva valimi hinnangud.

## 2.4 Kahe hetke disaini rakendus antud töös

Käesolevas töös valiti kattuvaks osaks kõik haiglad, sest haiglates töötab suurem osa tervishoiutöötajatest. Lisaks on teada, et haiglatel ei esine probleeme igal aastal andmeid esitada, kuna neil on selleks spetsiaalsed töötajad. Konstandi  $K$  arvutamine haiglate põhjal võtab arvesse iga-aastased muutused, aga ei võta arvesse väikeste TTOde muutusi. Mittekattuvad osad esimesel ja teisel hetkel jagati 2 aasta vahel, kus teise hetke valim on aasta, mille pära-

meetreid soovime leida ja esimese hetke valim on sellele eelnev aasta. Jagamine toimub samuti nagu veereva disaini korral ja on välja kirjutatud peatükis 3.2.

Antud töös katavad esimese ja teise hetke valimid kogu üldkogumi. Tähistame

$U_a$  – esimese hetke osa üldkogumist,

$U_m$  – esimese ja teise hetke kattuv osa,

$U_u$  – teise hetke mittekattuv osa üldkogumist. Eraldame nüüd esimese hetke mittekattuva osa

$$U_b = U_a \setminus U_m.$$

Kogu üldkogumi saab esitada järgmiste lõikumate osade ühendina  $U = U_b \cup U_m \cup U_u$ . Edasi vaatame kahte võimalust parameetrite hindamiseks.

### 2.4.1 Kahe hetke disain 1

Kirjutame välja teise hetke kogusumma:

$$t_y = \sum_U y_k = \sum_{U_b} y_k + \sum_{U_m} y_k + \sum_{U_u} y_k.$$

Siin on kaks viimast liidetavat teada ja on vaja hinnata ainult  $\sum_{U_b} y_k$ , mida saab teha  $y_k^0$  ja  $K$  abil. Antud hetke kogusumma hinnanguks on

$$\hat{t}_y = \sum_{U_b} y_k^0 + \sum_{U_m} y_k + \sum_{U_u} y_k, \quad (2.11)$$

kus  $y_k^0 = K z_k$  ja  $K$  hinnatakse kattuva osas võrrandist  $y_k = K z_k$  järgmiselt:

$$\begin{aligned} y_k &= K z_k, \quad \forall k \in U_m, \\ \sum_{k \in U_m} y_k &= \sum_{k \in U_m} K z_k = K \sum_{k \in U_m} z_k, \\ K &= \frac{\sum_{k \in U_m} y_k}{\sum_{k \in U_m} z_k}. \end{aligned}$$

Konstandi  $K$  väärtused tunnustele, mis on vaadeldud nii töö raames kui ka mainitud lisas 2, on välja toodud tabelis 4. Siin on kattuvaks osaks haiglate andmed. Tabelist nähtub, et paljude tunnuste korral on  $K$  ühe lähedane. Tervishoiutöötajate arvu korral on  $K$  alla 1,034, mis tähendab, et haiglates töötavate isikute arv suureneb aastatega vähem kui 3,4%. Samuti suureneb aastatega keskmine vanus, kuid mitte üle 3,5%. Tegelikult täidetud ametikoha puhul on  $K$  ühele veel lähedamal ja iga-aastane erinevus eelmisega ei ületa 2% ega lähe 2%-st madalamaks. Suuremad

aastased muutused on toimunud ülekoormusest tulenevate ametikohtade puhul, sest antud parameeter sõltub rohkem praegusest olukorrast kui eelmisest aastast.

Tabel 4. Kahe hetke disaini konstandi  $K$  väärtus igale vaadeldud parameetrile aastati

Aasta	Tervishoiutöötajate arv	Täidetud ametikohad tegeliku koormuse põhjal	Täidetud ametikohad tegeliku lepingulise põhjal	Ülekoormusest tulenevad ametikohad	Keskmine vanus
2014	1,0115	1,0223	1,0104	1,4001	1,0149
2015	1,0338	1,0076	0,9935	0,7577	1,0347
2016	1,0038	0,9818	0,9817	0,7472	1,0080
2017	1,0141	1,0057	1,0117	1,0371	1,0196
2018	1,0054	0,9950	1,0007	1,1140	1,0079

#### 2.4.2 Kahe hetke disain 2

Rakendame peatükis 2.3 sisse toodud tähistusi

$U = U_a \cup U_a^c$  üldkogum teisel hetkel,

$U_a = U_b \cup U_m$ , kus  $U_m$  on kattuv osa,

$U_a^c = U_u$  – mitte kattuv osa.

Olgu

$N$  – üldkogumi  $U$  maht (ridade arv),

$N_a$  – üldkogumi  $U_a$  maht,

$N - N_a$  – üldkogumi  $U_a^c$  maht,

$N_m$  – kattuva osa maht.

Kaasamistõenäosused on nüüd osakaalud:

$$\pi_{ak} = \frac{N_a}{N}; \quad \pi_{k|s_a} = \pi_{k|U_a} = \frac{N_m}{N_a};$$

$$\pi_{ak}^c = \frac{N - N_a}{N}; \quad \pi_{k|s_a^c} = \pi_{k|U_a^c} = \frac{N - N_a}{N - N_a} = 1.$$

Nüüd saame valemitest (2.9) ja (2.10) järgmised hinnangud:

$$\hat{t}_1 = \sum_{U_a} \frac{y_k^0}{N_a/N} + \sum_{U_m} \frac{y_k - y_k^0}{\frac{N_a}{N} \cdot \frac{N_m}{N_a}} =$$

$$= \frac{N}{N_a} \sum_{U_a} y_k^0 + \frac{N}{N_m} \sum_{U_m} (y_k - y_k^0),$$

$$\hat{t}_2 = \sum_{U_u} \frac{y_k}{\frac{N-N_a}{N}} = \frac{N}{N-N_a} \sum_{U_u} y_k.$$

Arvestades, et  $U_a = U_b \cup U_m$  ja  $U_b, U_m$  on lõikumatud ning  $K = \frac{\sum_{k \in U_m} y_k}{\sum_{k \in U_m} z_k}$ , saame

$$\begin{aligned} \hat{t}_1 &= \frac{N}{N_a} \sum_{U_a} K \cdot z_k + \frac{N}{N_m} \sum_{U_m} (y_k - K \cdot z_k) = \\ &= \frac{N}{N_a} \sum_{U_a} K \cdot z_k, \end{aligned}$$

millest

$$\begin{aligned} \hat{t}_1 &= \frac{N}{N_a} \cdot \left( \sum_{U_b} K z_k + \frac{\sum_{k \in U_m} y_k}{\sum_{k \in U_m} z_k} \sum_{U_m} z_k \right) = \\ &= \frac{N}{N_a} \cdot \left( \sum_{U_b} K z_k + \sum_{k \in U_m} y_k \right). \end{aligned}$$

Siin  $\hat{t}_1$  on  $U_a$  keskmine laiendatud üldkogumi  $U$  peale ja  $\hat{t}_2$  on  $U_u$  keskmine laiendatud üldkogumi peale.

Liithinnang  $\hat{t}_y = w_1 \hat{t}_1 + w_2 \hat{t}_2$  avaldub järgmiselt

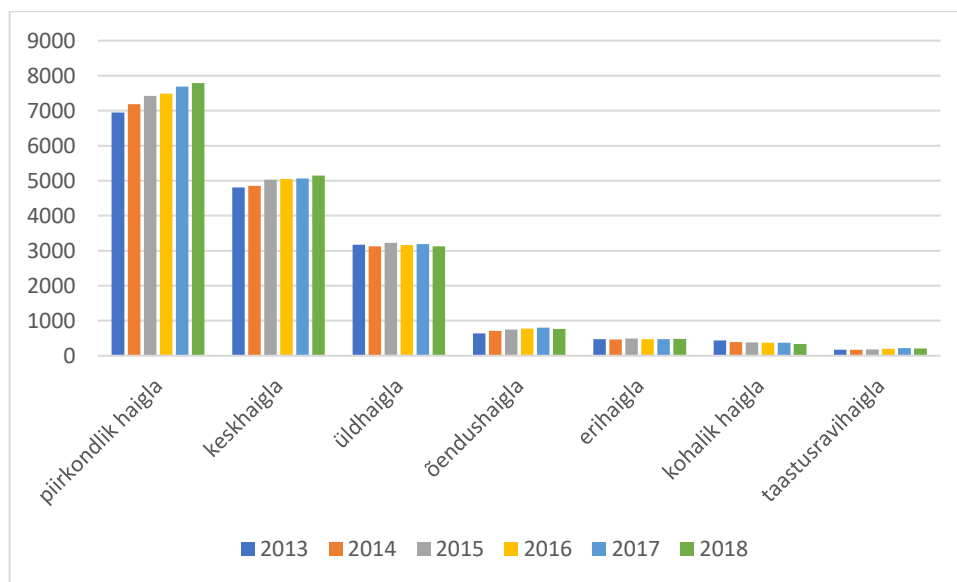
$$\hat{t}_y = w_1 \frac{N}{N_a} \cdot \left( \sum_{U_b} K z_k + \sum_{k \in U_m} y_k \right) + w_2 \frac{N}{N-N_a} \sum_{U_u} y_k. \quad (2.12)$$

### 3 Ettevalmistus praktiliseks ülesandeks

Magistritöös on kasutatud tervishoiutöötajate andmed aastate 2013–2018 kohta. Neid kasutades uurime, kui hästi väljapakutud meetodid lähendavad tegelikkust.

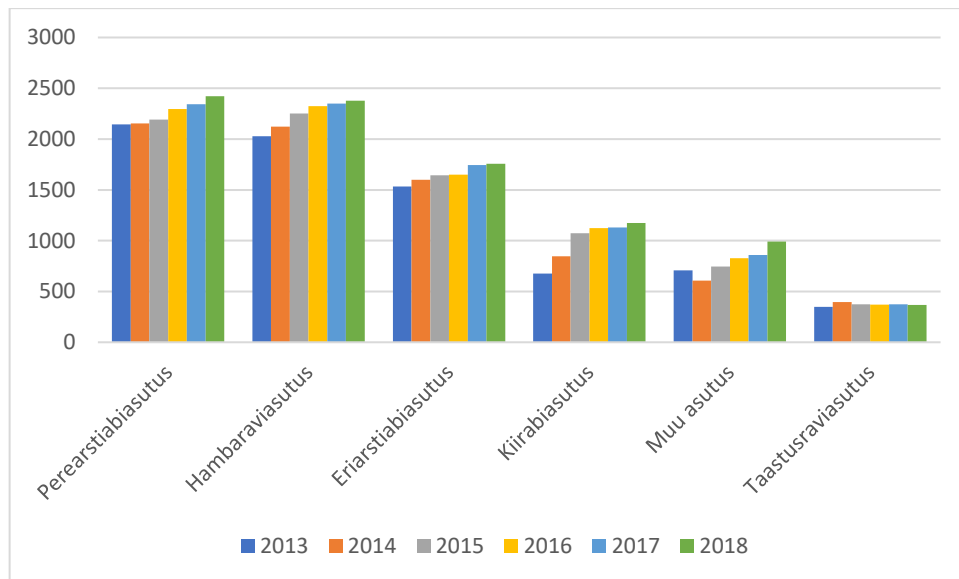
#### 3.1 Kihtide moodustamine

Üldkogumi osadeks jagamisel kasutatakse kihte selleks, et tagada võimalikult sarnased osad tervishoiuteenuse osutajate suhtes. Kuna aastate osad on struktuurilt sarnased, siis on ootuspäraselt hinnangud aastate lõikes väiksema varieeruvusega. Üldkogum on TTOde suhtes jagatud kihtideks järgmiselt: haiglad nende liikide järgi ja ülejäänud TTOd nende poolt osutavate tervishoiuteenuste järgi. Siinjuures on ühendatud teenused „üldarst (va perearst)“ ja „muu“ nende väikese mahu pärast. Gruppi „muu“ kuuluvad mujal nimetamata asutused, mis omavad tervishoiuteenuse osutamise tegevusluba, näiteks kaitsevägi ja koolitervishoid. Tervishoiutöötajate arv haigla liikide ja aastate lõikes on näha joonisel 1. Erihaiglad on sellised haiglad, mis ei kuulu haiglavõrgu arengukava haiglate hulka ja kus tehtav töö on keskendunud mõnele kitsamale erialale, nt sünnitusabi ja günekoloogia. Jooniselt on näha, et oluliselt rohkem tervishoiutöötajaid on piirkondlikes, kesk- ja üldhaiglates. Ülejäänud haiglates on tervishoiutöötajate arv alla tuhande.



Joonis 1. Tervishoiutöötajate arv haigla liikide ja aasta lõikes

Mittehaiglad on jagatud TTO poolt osutavate teenuse liikide järgi. Joonisel 2 on näha, et erinevate teenustega asutustes töötavate isikute arv ei ületa 2500.



Joonis 2. Tervishoiutöötajate arv TTO teenuse (v.a haiglad) ja aasta lõikes

### 3.2 Algoritm üldkogumi jagamiseks kaheks osaks

Aruandluse koormuse vähendamiseks jagati TTOde üldkogum kaheks osaks (kaheks pooleks). Esimeses pooles olevatelt TTOdelt (asutustelt) küsiti aruandeid paaritutel aastatel (2013, 2015 jne) ja teises pooles olevatelt TTOdelt paarisaastatel (2014, 2016 jne). Jagamist rakendati igas kihis eraldi.

Igas kihis olevatele asutustele arvutati TTO suurus (näiteks tervishoiutöötajate arv), seejärel järjestati TTOd selle põhjal.

Kihis  $h$  rakendati järgmist jagamise algoritmi (algoritmi läbimisel on oluline järgida alltoodud järjekorda):

0. Olgu kihis  $h$   $M_h$  asutust, mis on järjestatud TTO suuruse järgi. Kihhi loend on  $\{1, 2, \dots, M_h\}$ , kus TTO  $M_h$  on kihis suurim.
1. Viimane asutus (suurim asutus) läheb 1. poolde.
2. Kui ülejäänud asutuste töötajate kogusumma on väiksem kui viimase asutuse töötajate arv, siis kõik ülejäänud lähevad 2. poolde. Ja algoritm lõpeb. Vastasel juhul algoritm jätkub sammuga 3.
3. Kui kihis on kuni 4 asutust, siis jagatakse asutused järgmiselt: esimene ja viimane lähevad 1. poolde ning ülejäänud 2. poolde. Ja algoritm lõpeb. Vastasel juhul algoritm jätkub sammuga 4.

4. Kui kihis on rohkem kui 4 asutust, siis esimene asutus läheb 1. poolde (nüüd esimene ja viimane on esimeses pooles, teised on jagamata).
5. Teise poolde lähevad tagantpoolt  $j$  asutust  $(M_h - 1), \dots, (M_h - j)$ , kus  $j$  on suurim indeks, mille korral kehtib võrratus

$$TöötajateArv_1 + TöötajateArv_{M_h} \geq \sum_{i=1}^j TöötajateArv_{M_h-i},$$

kus  $TöötajateArv_i$  on töötajate arv asutuses  $i$ .

6. Teine asutus läheb 2. poolde.
7. Asutus 3 ja  $(M_h - (j - 1))$  lähevad 1. poolde.
8. Nüüd määrame asutustele indeksid  $N$  ja  $k$ . Indeksi  $N$  määramiseks vaatleme tagantpoolt asutusi, millele on juba pool määratud. Nendest tähistame indeksiga  $N$  asutuse, mis on tagantpoolt viimane. Indeksi  $k$  määramiseks vaatleme eespoolt asutusi, millele pole poolt määratud ning indeksiga  $k$  tähistame nende hulgast esimese asutuse. Pärast 7. sammu  $N = (M_h - (j - 1))$  ja  $k = 4$ .
9. Nüüd 5. sammuga sarnaselt valitakse, kui paljud ülejäänud suurematest asutustest lähevad teise poolde. Sinna lähevad asutused  $(N - 1), \dots, (N - j)$ , kus  $j$  on võimalikult suur indeks mille korral kehtib  $(N - j) > k$  ja

$$\sum_{\substack{\text{asutus } i \text{ on} \\ \text{1. pooles}}} TöötajateArv_i \geq \sum_{\substack{\text{asutus } i \text{ on} \\ \text{2. pooles}}} TöötajateArv_i + \sum_{i=1}^j TöötajateArv_{N-i}.$$

10. Asutus  $k$  läheb 2. poolde.
11. Asutus  $k + 1$  ja  $(N - (j - 1))$  lähevad 1. poolde.
12. Algoritm korratakse punktist 8 kuni kõik asutused on jagatud.

Antud algoritm on realiseeritud tarkvaras R. Vastav kood on toodud Lisas 1.

### 3.3 Aastatega muutuv andmestik

Eespool kirjeldatud olukord sobib siis, kui üldkogum aastatega ei muutu. Praktikas aga toimuvad üldkogumis pidevad muutused. Aja jooksul tekivad uued TTOd ja mõned suletakse. Praktikas kasutatakse nn tahapoolse valikut (*back samples*), valides uued asutused mitte ainult praeguse, vaid ka eelneva valimi hulka [11].

Käesolevas magistritöös kasutati järgmist eeskirja uute tervishoiuasutuste kaasamiseks vaadeldaval hetkel ja küsitlemisaasta määramiseks. Igal aastal kontrollitakse kõikide tervishoiuasutuste reaalset tegutsemist vaadeldaval aastal. Kui asutus tegutses, siis ta jääb üldkogu-

misse ja sellele varem määratud poolde. Kui asutus ei tegutsenud, siis uuendatud üldkogumist asutus eemaldatakse, kuid enne määratakse kiht ja pool, kus asutus oli ning palju töötajad tal oli eelmisel aastal. See aitab määrata uue asutuse asukohta pooltes.

Kui tulid juurde uued asutused, mida varem andmestikus polnud, siis nende kohta on teada ainult kiht. Vaja on valida millisesse poolde nad lähevad. Olgu  $U_{us}$  uute asutuste arv,  $K1$  ja  $K2$  on vastavalt esimesest ja teisest poolest kustutatud asutuste arv ja  $KT1$  ja  $KT2$  on vastavalt esimesest ja teisest poolest kustutatud asutuste suuruste kogusumma. Uute asutuste arv, mis lähevad esimesse poolde ( $U_{us1}$ ) leitakse järgmiselt

$$U_{us1} = U_{us} \cdot \frac{1}{2} \left( \frac{K1}{K1 + K2} + \frac{KT1}{KT1 + KT2} \right).$$

Uute asutuste hulgast valitakse juhuslikult  $U_{us1}$  asutust, mis lähevad 1. poolde ja ülejäänud  $U_{us} - U_{us1}$  asutust lähevad 2. poolde. Antud eeskirja rakendatakse igas kihis ja igal aastal eraldi.

### 3.4 Moodustatud valimite ülevaade

Rakendades peatükis 3.2 toodud algoritmi, on kõik asutused jagatud oma pooltesse, mida kirjeldab täpsemalt tabel 5. Käesolevas töös kasutati teadaolevaks baasiks (loendusandmeteks) 2013. aasta andmeid. Eraldi vaadeldi kahte võimalust – kas jagada asutused töötajate või objektide (isik-amet) arvu põhjal. Töötajad on unikaalsed, aga objektid on tabeli read, kus isik on kordusega, kui ta on erineval ametikohal või erinevas asutuses. Mõlema jagamisviisi korral jäi asutuste arv kihtides samaks. Kihtides olevate asutuste arvud esimeses ja teises pooles on toodud tabelis 5. Üldkokkuvõttes on esimeses pooles 665 ja teises 696 asutust. Võrreldes esimese ja teise poole töötajate ning objektide arvusid tabelis 5, siis need on mõnevõrra erinevad. Aga kui vaadata osakaalusid, siis nii esimese kui teise poole osakaal on ligikaudu 0,5. Näiteks on erihaiglas esimeses pooles 215 ja teises 225 töötajat, mis tähendab esimese poole osakaalu  $\frac{215}{215+225} \approx 0,5$  ja vastavalt ka teise poole osakaalu  $\approx 0,5$ . Objekte on esimeses pooles 253 ja teises 229, mistõttu esimese poole osakaal ja vastavalt ka teise poole oma on 0,5. Seega jagamise tulemus töötaja- ja objekti-põhiselt on sarnane.

Veereva disaini 2 ja kahe hetke disaini jaoks võeti kattuvaks osaks kõik haiglad (kihid 11 – 17), pooled on antud juhul samad, mis hetked. Kattuval osal olevaid asutusi küsitletakse igal aastal. Ülejäänud TTOd jagunesid nii nagu tabelis 5 alates kihist 20.



Tabel 5. Kihid kahe valimiga veereva disaini 1 korral

Kiht	Kihi nimi	Kihi maht	Esimeses pooles	Teises pooles	Esimese poole asutuse töötajate ja objektide arv		Teise poole asutuse töötajate ja objektide arv	
11	Piirkondlik haigla	3	1	2	3533	3694	3451	3467
12	Keskhaigla	4	2	2	2665	2701	2223	2224
13	Üldhaigla	11	5	6	1760	1805	1452	1525
14	Taastusravihaigla	4	1	3	100	100	76	79
15	Õendushaigla	27	13	14	328	329	311	315
16	Erihaigla	10	5	5	251	253	228	229
17	Kohalik haigla	4	2	2	215	221	225	228
20	Perearstiabi	478	238	240	1103	1112	1101	1111
25	Eriarstiabi	238	117	121	814	822	815	823
30	Hambaravi	468	233	235	1117	1132	1116	1131
40	Kiirabi	5	2	3	360	363	324	325
50	Taastusravi	30	13	17	177	179	177	181
51	Diagnostika	15	3	12	83	83	83	83
53	Õendusabi	42	19	23	106	107	107	108
100	Ülejäänud (Muu ja üldarst, v.a perearst)	22	11	11	177	177	164	165
	<b>Kokku</b>	<b>1361</b>	<b>665</b>	<b>696</b>		<b>13078</b>		<b>11994</b>

Tabelis 6 näeme, kuidas kasutatud algoritmi tulemusel jaotusid TTOd pooleks maakonniti kahe erineva jagamise viisi korral. Väikesed muutused on toimunud sõltuvalt jagamisviisist. Näiteks Harjumaal jagunesid asutused samamoodi mõlema jagamisviisi korral, aga teiste maakondade puhul veidi erinevalt.

Tabel 6. TTOde pooleks jaotumine maakonniti

	Esimese poole asutuste arv		Teise poole asutuste arv		Kokku maakonnas
Jagamisviisid	Töötaja-põhine	Objekti-põhine	Töötaja-põhine	Objekti-põhine	
Harju maakond	44	44	46	46	90
Hiiu maakond	5	6	8	7	13
Ida-Viru maakond	58	62	68	64	126
Jõgeva maakond	20	19	21	22	41
Järva maakond	14	16	18	16	32
Lääne maakond	12	14	16	14	28
Lääne-Viru maakond	39	43	37	33	76
Põlva maakond	16	19	20	17	36
Pärnu maakond	38	35	41	44	79
Rapla maakond	13	12	16	17	29
Saare maakond	25	24	15	16	40
Tartu maakond	18	15	26	29	44
Valga maakond	16	17	16	15	32
Viljandi maakond	30	36	32	26	62
Võru maakond	19	18	23	24	42
Tallinn	238	223	223	238	461
Tartu linn	60	62	70	68	130
Kokku aastas	<b>665</b>		<b>696</b>		<b>1361</b>

## 4 Saadud tulemuste analüüs

Kokku vaadeldakse töös 4 meetodit:

1. Veerev disain 1 on kahe perioodiga veerev disain, mis rakendatakse kogu üldkogumile.
2. Veerev disain 2 on osaline veerev disain, kus veeremine toimub mittehaiglate hulgas.
3. Kahe hetke disain 1 korral arvutatakse hinnang otseselt üldkogumi kogusumma avaldisest. Kasutatakse prognoosimiskonstanti  $K$ .
4. Kahe hetke disain 2 korral tuletatakse hinnang kahe hetke disaini hinnangust, kus valimid on asendunud üldkogumi osadega ja kaasamistõenäosused on asendunud vastavate osakaaludega.

Lisaks vaadeldakse iga meetodit kahe jagamisviisi korral, st kas üldkogum on aastate vahel osadeks jagatud töötaja- või objekti-põhiselt (vt. tabel 5).

Kokku uuriti 6 aastat: 2013 – 2018, kus aasta 2013 oli baas, mida kasutati küsitletud asutuste pooleks jagamiseks kahe erineva jagamisviisiga. Ülejäänud aastate puhul leiakse hinnangud kasutades vaadeldavat aastat ja sellele eelnevat aastat. Kokku vaadeldi 5 paari: 2013–2014, 2014–2015, 2015–2016, 2016–2017, 2017–2018.

### 4.1 Tervishoiutöötajate arvu hinnangud

Esmalt uurime tervishoiutöötajate koguarvu hinnanguid veereva disaini korral. Märgime, et andmestikus ilmub isik mitu korda, kui ta töötab mitmes asutuses või mitmel ametikohal. Tervishoiutöötajate koguarvuks on vaja leida unikaalsete isikute arv. Kuna TToD, kus isik töötab, saavad sattuda veereva disaini erinevasse poolde, siis unikaalsete isikute väljaselgitamiseks on vaja vaadelda mõlemaid aastaid koos. Kasutame hinnangute leidmiseks peatükis 2.2.1 leitud valemit (2.1). Väärtused tuleb leida ka kaaludele, oletades, et eelmise aasta andmed on vähemtähtsad käesoleva aasta jaoks. Sobivaima kaalu selgitasime katseliselt, vaadates erinevaid kaale: isikud, kes esinesid andmestikus vaadeldaval aastal, on kaaluga (%)  $w = 50, 60, 70, 80, 90$  ja isikud, kes esinesid ainult eelneval aastal on kaaluga  $100 - w$ . Aastal 2013 võeti hinnangute asemel tegelikud väärtused, sest sellele aastale eelnevat aastat ei olnud.

Veereva disaini 2 korral arvutatakse eraldi haiglates töötavate isikute arv  $M_h$  ja valemiga (2.4) hinnatakse nende arv, kes ei tööta haiglates. Lõpphinnang tervishoiutöötajate arvule saadakse valemist (2.5).

Mõlema veereva disaini hinnanguid vaadeldi erinevates andmestikes, mis tekkisid vastavalt jagamise viisidele. Esimesel viisil jagati asutused unikaalsete töötajate ehk isikute järgi ja teisel objektide ehk tabeli ridade (isik-amet TTOs) järgi.

Selleks, et võrrelda hinnanguid erinevate jagamisviiside ja disainimeetodite korral, kasutame lähendusviga, mis on ruutjuur keskmisest ruutveast,

$$RMSE = \sqrt{\sum_{i=2014}^{2018} (t_i - \hat{t}_i)^2 / 5}, \quad (4.1)$$

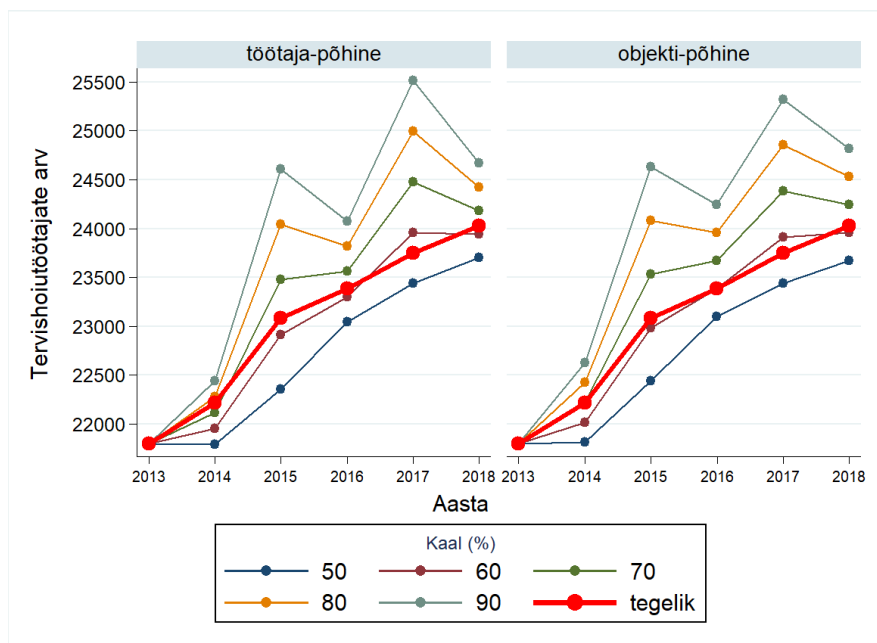
kus  $t_i$  on aasta  $i$  tegelik väärtus ja  $\hat{t}_i$  hinnang. Mida väiksem on  $RMSE$  väärtus, seda täpsem, seega ka eelistatavam, on hinnang.

Tervishoiutöötajate arvu hinnangute  $RMSE$  on toodud tabelis 7. Sealt näeme, et täpseimad veereva disaini 1 ja 2 hinnangud mõlema jagamisviisi korral on kaaluga 60. Samuti näeme, et jagamine objektide arvu järgi annab ebatäpsema tulemuse, sest  $RMSE$  on suurem. Näeme, et kõik veereva disaini 2 lähendusvead on väiksemad kui veereva disaini 1 korral. See tähendab, et haiglate pärisandmete kasutamine igal aastal aitab teha hinnangud täpsemaks, sest nüüd hinnatakse ainult mittehaiglate aastast muutust.

Tabel 7. Tervishoiutöötajate arvu hinnangute lähendusvea  $RMSE$  väärtus erinevate kaalude ja jagamise viisi lõikes veereva disaini 1 ja 2 korral

Meetod	Veerev disain 1		Veerev disain 2	
Kaal $w$ (%)	Töötaja-põhine jagamine	Objekti-põhine jagamine	Töötaja-põhine jagamine	Objekti-põhine jagamine
50	368,8	368,8	142,4	147,4
60	93,9	112,2	67,6	66,7
70	342,4	354,6	88,7	115,5
80	698,0	716,2	143,5	182,1
90	1053,6	1077,9	198,2	248,8

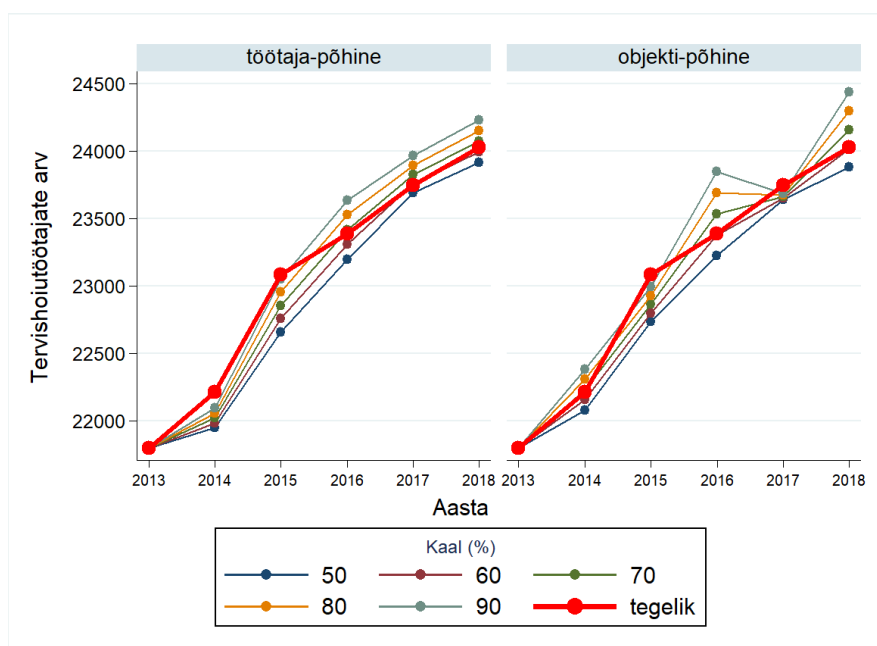
Joonisel 3 on näidatud tervishoiutöötajate koguarvu tegelik väärtus (punane joon) ja veereva disaini 1 hinnangud erinevate kaaludega, kus jagamisviis on kas töötajate või objektide arvu põhine. Hinnangud kaaludega 50 on stabiilsed, st ilma suurte hüpeteta, kuid nad alahindavad tegelikkust. Mida suurem on kaal, seda rohkem ülehinnatakse koguarvu. Vastavalt jagamisalgoritmile on suurim asutus esimeses pooles (paaritud aastad). Seega nendel aastatel toimub hinnangute kõikumine üles. Joonise põhjal näeme ka, et kaal 60 sobib kõige paremini töötajate arvu hindamiseks, kuna tulemused peegeldavad tegelikkust kõige paremini.



Joonis 3. Tervishoiutöötajate arvu hinnangud veereva disaini 1 korral jagamise viisi lõikes

Veereva disaini 2 hinnangud järgivad tegelikkust kõikide kaalude korral paremini (joonis 4).

Haiglate andmete iga-aastaselt kasutamisel on hinnangule suur stabiliseeriv mõju.



Joonis 4. Tervishoiutöötajate arvu hinnangud veereva disaini 2 korral jagamise viisi lõikes

Kahe hetke disaini 1 korral arvutati tervishoiutöötajate arv valemi (2.11) abil, kus tunnused  $y_k$  ja  $z_k$  on nüüd unikaalsuse indikaatorid  $I_{d-unik}(k)$  (vt valem (1.3)), milles tingimus  $d$  viitab kas vaadeldavale või eelnevale aastale. Kattuvaks osaks on kõik haiglad ja konstandi  $K$  väärtus on toodud tabelis 4.

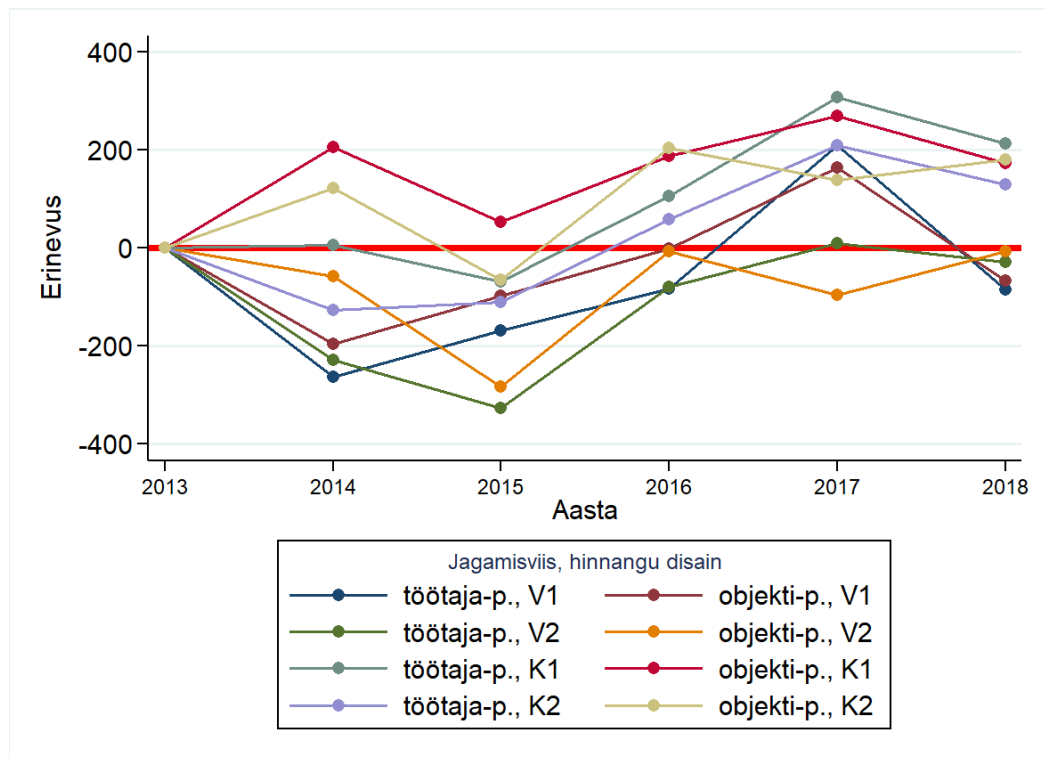
Kahe hetke disaini 1 hinnang ei kasuta kaale. Selle disainiga saadud hinnangute lähendusvead on 140,3, kui asutuste jagamine on töötaja-põhine ja 177,8 kui jagamine on objekti-põhine.

Viimasena vaadati kahe hetke disaini 2 hinnanguid (valem (2.12)), kus kaaludena (%) kasutati  $w_2 = 10, 20, 30, 40, 50, 60, 70, 80, 90$  ja  $w_1 = 100 - w_2$  ning konstandi  $K$  väärtus on sama, mis kahe hetke disaini 1 korral (tabel 4). Saadud hinnangute lähendusvigade väärtused on toodud tabelis 8 ja graafik lisas 2. Tabelis 8 on näha, et kõige täpsemad hinnangud on kaaluga 20 ja kaalude suurendamisel muutuvad hinnangud ebatäpsemaks. Märgime, et  $w_2$  on vaadeldavate aastate väiksemate asutuste (mittehaiglate) kaal.

Tabel 8. Lähendusvea väärtus erinevate kaalude ja jagamise viisi lõikes kahe hetke disainide korral

Meetod	Kaalud $w_2$ (%)	Töötaja-põhine jagamine	Objekti-põhine jagamine
<b>Kahe hetke disain 1</b>		140,3	177,8
<b>Kahe hetke disain 2</b>	10	257,9	303,2
	20	127,2	142,3
	30	182,1	246,4
	40	396,1	455,3
	50	610,1	664,1
	60	824,1	872,9
	70	1038,2	1081,7
	80	1252,2	1290,5
	90	1466,2	1499,4

Ülal kirjeldatud nelja meetodi abil saadud hinnangutest valime igast täpseima (vähima lähendusveaga) hinnangu. Joonisel 5 on toodud nende hinnangute erinevus tegelikust väärtusest. Meetodid on kodeeritud järgmiselt: V1 – Veerev disain 1, V2 – Veerev disain 2, K1 – Kahe hetke disain 1, K2 – Kahe hetke disain 2. Kuna vaadatakse ka kahte jagamisviisi (töötajad, objektid), siis on hinnanguid kokku 8. Sealt on näha, et mõned hinnangud, mis hindavad täpsemalt alguses, muutuvad kehvemaks viimastel aastatel ja vastupidi hinnangud, mis oli kehvad uuritava perioodi alguses, muutuvad aastatega paremaks.



Joonis 5. Tervishoiutöötajate arvu parimate hinnangute erinevus tegelikkust iga meetodi korral

Arvestades nii lähendusviga kui ka erinevuste joonist, tunnistame parimaks meetodiks veereva disaini 2. Joonisel näeme, et viimasel kolmel aastal on hinnangud kõige lähemal tegelikule väärtusele. Ka üle kõigi aastate on sellel meetodil väikseim lähendusviga (tabel 7, kaal 60). Saame järeldada, et jagamisviis ei mõjuta antud meetodit palju.

## 4.2 Tegelikult täidetud ametikohtade hinnangud

Tervishoiutöötajate töökoormuse kirjeldamiseks on kasutusel kokku kolm tunnust: täidetud ametikohad lepingulise koormuse järgi, tegelikult täidetud ametikohad ja ületundidest tulenev täiendav täidetud ametikohtade arv. Kõigi hinnangute joonised ja lähendusvigade tabel on toodud lisas 2. Antud peatükis vaatame lähemalt tegelikult täidetud ametikohtade hinnangut.

Veereva disaini 1 ja 2 hinnangutes kasutati samasuguseid kaalusid, nagu tervishoiutöötajate arvu hinnangu leidmiseks: isikud, kes esinesid andmestikus vaadeldaval aastal, on kaaluga (%)  $w = 50, 60, 70, 80, 90$  ja isikud, kes esinesid ainult eelneval aastal on kaaluga  $100 - w$ . Lisaks kasutati kaalumist tervishoiutöötajate osakaaludega (valem (2.2)).

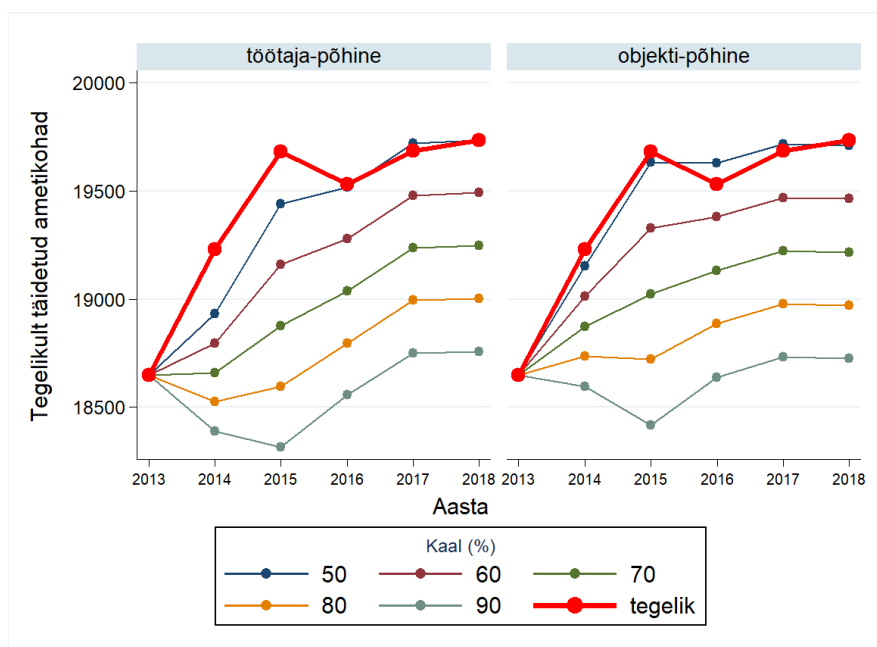
Hinnangute lähendusvead ( $RMSE$  valem (4.1)) veerevate disainide korral on toodud tabelis 9. Sealt on näha, et aastale vastav osakaalu suurendamisel muutub hinnang ebatäpsemaks ja pari-

ma hinnangu saame kaaluga 50. See tähendab, et tervishoiutöötajate arvu kasutamine aitas hinnangute kaalumisel ja teist kaalumist ei ole vaja. Veereva disaini 2 puhul on lähendusvigade väärtused väiksemad kui veereva disaini 1 korral.

Tabel 9. Tegelikult täidetud ametikohtade hinnangu lähendusvea väärtus veereva disaini 1 ja 2 korral erinevate kaalude ja jagamise viisi lõikes

Meetod	Veerev disain 1		Veerev disain 2	
	Töötaja-põhine jagamine	Objekti-põhine jagamine	Töötaja-põhine jagamine	Objekti-põhine jagamine
50	117,6	55,5	193,1	102,5
60	332,0	241,3	234,5	148,5
70	560,8	477,8	275,8	194,5
80	789,5	714,3	317,2	240,5
90	1018,2	950,9	358,5	286,5

Joonisel 5 on näha, et saadud hinnangud on stabiilsemad üles-alla liikumise osas, kui olid tervishoiutöötajate arvu hinnangud samal meetodil (joonis 3). Kui kaalud on viiekümnest suuremad, siis toimub alahinnang. See kinnitab, et tehtud kaalumine töötajate arvu hinnanguga aitas teiste tunnuse hindamisel ja lisakaalumine on üleliigne. Viimane viib alahinnangutele. Hea tulemuse saamisele aitab kaasa see, et koormuse tunnuse kogusumma leidmine on lihtsam kui töötajate arvu oma, sest ei ole vaja leida unikaalseid isikuid.

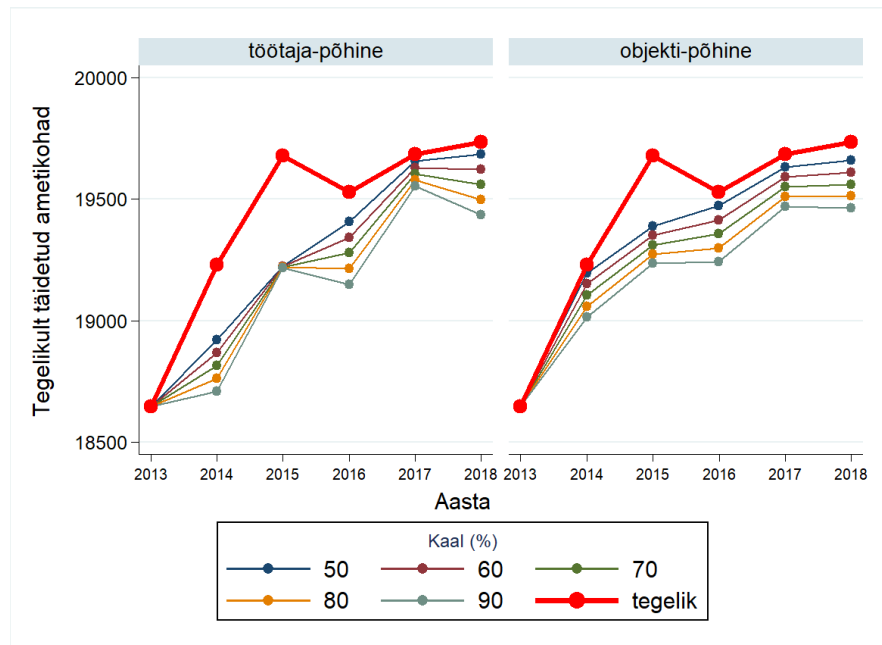


Joonis 6. Tegelikult täidetud ametikohtade hinnangud veereva disaini 1 korral jagamise viisi lõikes

Veereva disaini 2 hinnangud on toodud joonisel 7. Siin on hinnangud väiksema varieeruvusega kui veereva disaini 1 korral, aga aastal 2015 toimunud hüpet ei suudeta piisavalt hästi kajastada.



Põhjuseks on see, et asutusi, kus toimusid suured muutused ei küsitletud 2015. aastal. Samal joonisel on näha kummaline olukord, kus aastal 2015 läbivad töötaja-põhise jagamisviisiga hinnangud kõikide kaalude korral ühte punkti. Põhjus selgub valemit (2.7) uurides. Selle liidetakse on enne kaalumist praktiliselt võrdsed:  $\frac{\hat{M}_{MH}}{\hat{M}_{0-MH}} t_{y_{MH-0}} = 5555$  ja  $\frac{\hat{M}_{MH}}{\hat{M}_{1-MH}} t_{y_{MH-1}} = 5548$ . Summeerides neid numbreid mistahes kaaludega (kaalud summeeruvad üheks), tulemused praktiliselt ei muutu.



Joonis 7. Tegelikult täidetud ametikohtade hinnangud veereva disaini 2 korral jagamise viisi lõikes

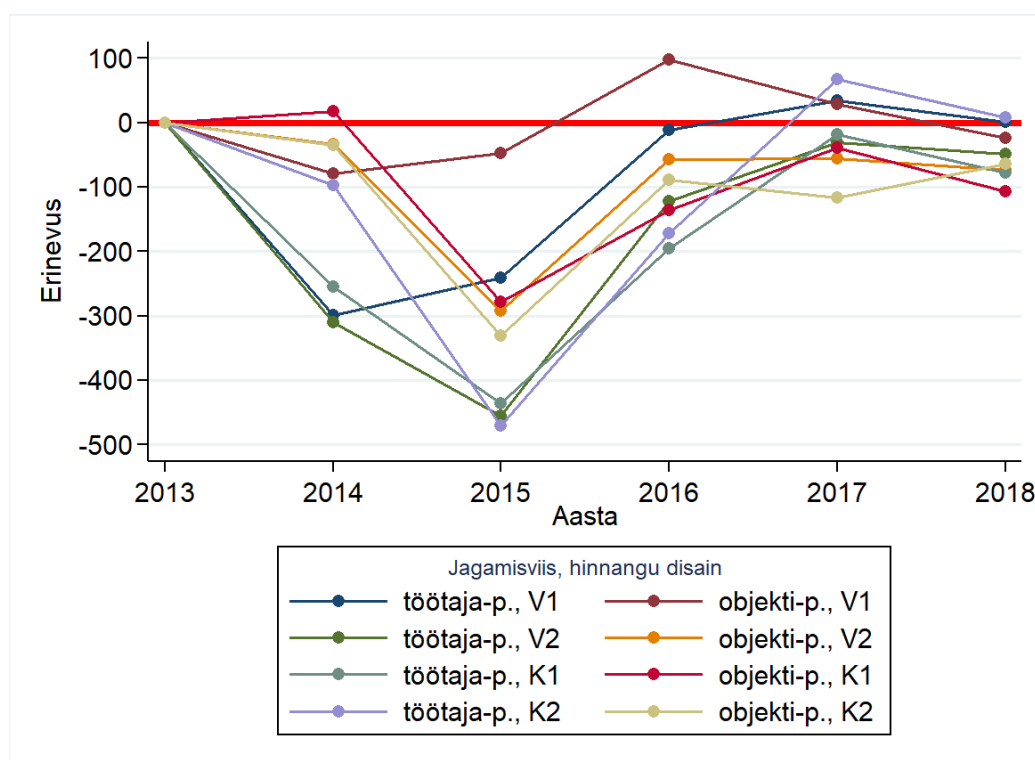
Kahe hetke disaini hinnangute arvutamisel oli kaks viisi. Esimene on kogusumma avaldise kaudu (valem (2.11)) ja teine disaini kaudu (valem (2.12)). Esimesel viisil kasutati teise hetke kogusummat ja eelmise hetke konstandiga  $K$  kaalutud kogusummat ning teisel viisil kasutati kaasamistõenäosusest tuletatud kaalusid  $\frac{N}{N_a}$  ja  $\frac{N}{N-N_a}$  ning kaalusid  $w_1$  ja  $w_2$ , mis on teadmata. Vaadeldi kaale (%)  $w_1 = 10, 20, 30, 40, 50, 60, 70, 80, 90$  ja  $w_2 = 100 - w_1$ . Konstandi  $K$  väärtus on toodud tabelis 4.

Tabelis 10 on näha lähendusvea väärtused mõlema kahe hetke disaini korral. Kahe hetke disaini 2 puhul on parimad hinnangud kaaluga 10 töötajate arvu põhise jagamise viisi puhul ja kaaluga 20 objektide arvu põhise jagamise korral. Kahe hetke disaini 1 hinnangud on sama täpsed kui kahe hetke disaini 2 omad.

Tabel 10. Tegelikult täidetud ametikohtade hinnangu lähendusvead kahe hetke disaini 1 ja 2 korral erinevate kaalude ja jagamise viisi lõikes

Meetod	Kaalud $w_2$ (%)	Töötaja-põhine jagamine	Objekti-põhine jagamine
<b>Kahe hetke disain 1</b>		196,7	115,5
<b>Kahe hetke disain 2</b>	10	163,2	169,9
	20	233,0	127,6
	30	333,3	218,8
	40	433,5	371,4
	50	533,8	524,0
	60	634,1	676,7
	70	734,4	829,2
	80	848,5	981,8
	90	977,7	1134,5

Joonisel 8 on näha iga meetodi parimate hinnangute erinevus tegelikust väärtusest. On näha, et enamuses neist hinnatakse tegelik väärtus alla. Viimasel kahel aastal näitavad kõik hinnangud head tulemust. Kuid ainult üks hinnangutest näitas head tulemust igal aastal. See on objekti-põhise jagamisviisiga veereva disaini 1 hinnang. Sama tulemust näeme ja ka tabelis 9 ja joonisel 6.



Joonis 8. Tegelikult täidetud ametikohtade parimate hinnangute erinevus tegelikkusest iga meetodi korral

TAI avaldab ka teisi tervishoiutöötajate koormust iseloomustavaid hinnanguid: täidetud ametikohad lepingulise koormuse järgi ja ületundidest tulenev täiendav täidetud ametikohtade arv.

Neid ei ole käesoleva töö raames põhjalikumalt analüüsitud, graafikud ja lähendusvead on välja toodud lisas 2.

### 4.3 Hinnangud keskmisele vanusele

Keskmise vanuse leidmiseks on vaja leida unikaalsete isikute vanuse kogusumma ja unikaalsete isikute koguarv. Nende suhe on keskmine vanus. Unikaalsete isikute arvu hinnangud on toodud peatükis 4.1. Vanuse kogusumma hinnangud leiti järgmiste valemitega: (2.3) veereva disaini 1 korral, (2.7) veereva disaini 2 korral ning (2.11) ja (2.12) vastavalt kahe hetke disaini 1 ja 2 korral.

Hinnangute lähendusvead veerevate disainide korral on toodud tabelis 11. On näha, et veerev disain 1 hindab kehvemini kui veerev disain 2. Veereva disaini 1 parim hinnang saavutatakse kaaluga 80 mõlema jagamisviisi korral. Veereva disaini 2 korral on kõik hinnangute vead alla ühe vanuseaasta ja parimad on kaaluga 70, kus viga ei ole suurem kui 0.2 vanuseaastat.

Tabel 11. Keskmise vanuse hinnangute lähendusvead veereva disaini 1 ja 2 korral erinevate kaalude ja jagamise viisi lõikes

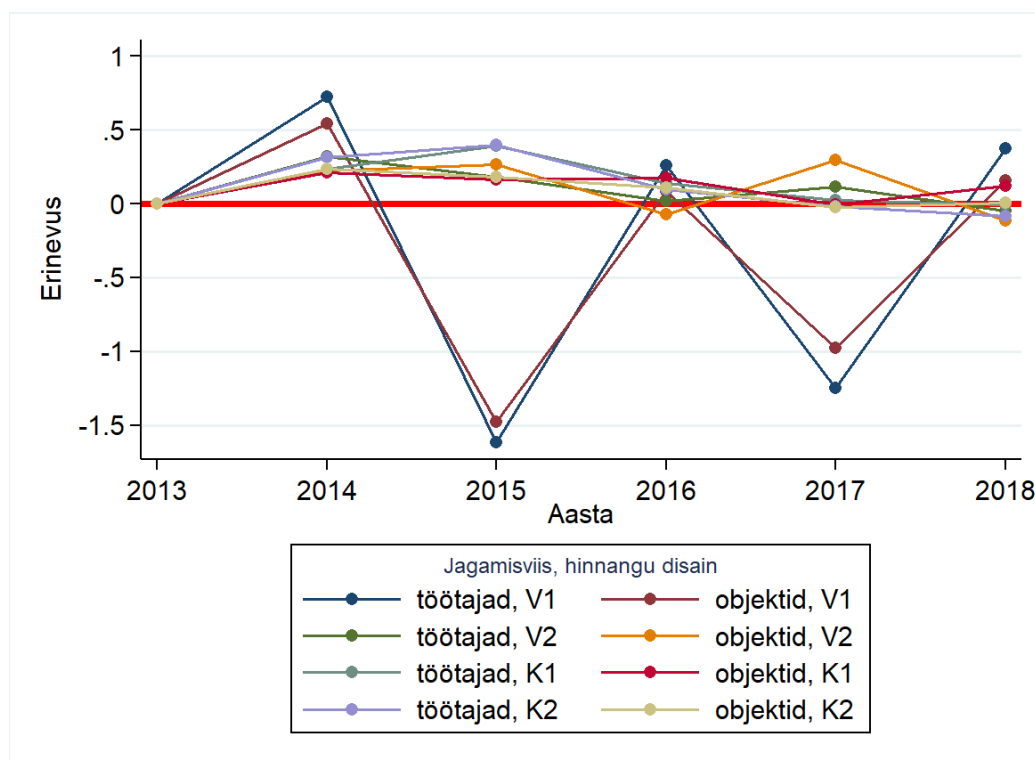
Meetod	Veerev disain 1		Veerev disain 2	
Kaal w (%)	Töötaja-põhine jagamine	Objekti-põhine jagamine	Töötaja-põhine jagamine	Objekti-põhine jagamine
50	3,40	3,48	0,64	0,71
60	2,12	2,16	0,38	0,41
70	0,89	0,89	0,14	0,20
80	0,85	0,65	0,20	0,26
90	1,46	1,53	0,40	0,47

Kahe hetke disaini 1 ja 2 lähendusvead on toodud tabelis 12. On näha, et kõik neist on paremad kui veereva disaini 1 hinnangud. Kahe hetke disaini 2 parimad hinnangud on kaaluga 20 ja nad on sama täpsed kui kahe hetke disaini 1 hinnangud.

Tabel 12. Lähendusvea väärtus keskmisele vanusele kahe hetke disaini 1 ja 2 hinnangule kaalude ja jagamise viisi lõikes

Meetod	Kaalud $w_2$ (%)	Töötaja-põhine jagamine	Objekti-põhine jagamine
<b>Kahe hetke disain 1</b>		0,16	0,14
<b>Kahe hetke disain 2</b>	10	0,20	0,18
	20	0,18	0,11
	30	0,28	0,18
	40	0,42	0,29
	50	0,56	0,41
	60	0,71	0,52
	70	0,85	0,64
	80	1,01	0,76
	90	1,16	0,88

Keskmise vanuse parimad hinnangud on toodud joonisel 9, kus vaadeldakse erinevust tegelikust väärtusest. Siin on näha, et veerev disain 1 andis ebasoovitava tulemuse, sest paaritutel aastatel tekib alahinnang. See viitab asjaolule, et asutuste jagamisel eri aastate vahel sattusid esimesse poolde nooremad inimesed kui teise poolde. Haiglate andmete iga-aastaselt kasutamisel seda mõju enam ei ole.



Joonis 9. Keskmise vanuse parimate hinnangute erinevus tegelikkusest iga meetodi korral

Keskmise vanuse korral ei saa eelistada ühte meetodit teisele, kui välistada veerev disain 1. Objekti-põhise jagamise korral, andsid nii veereva disaini 2 kui ka mõlemad kahe hetke disainiga tehtud hinnangud sama hea tulemuse.

## Kokkuvõte

Käesolevas magistritöös uuriti, milline neljast meetodist sobib tervishoiutöötajate aruandluskoormuse vähendamiseks. Vaadeldi veerevat ja kahe hetke disaini, millest kumbagi rakendati üldkogumis kahel erineval viisil. Veerevat disaini 1 rakendati kogu üldkogumis, veerevat disaini 2 aga üksnes üldkogumi osas. Kahe hetke disaini 1 ja 2 hinnangud avaldati kahel erineval meetodil. Üldiselt avaldusid hinnangud kaalutud lineaarkombinatsiooni kujul. Erinevate kaalude mõju uuriti katseliselt.

Aruandluskoormuse vähendamiseks kasutati andmeesitajate ehk tervishoiuasutuste kaasamist erinevatel aastatel. Osa tervishoiuasutustest otsustati küsitleda paaritutel aastatel ja teine osa paarisaastatel. Selleks, et valida, milline asutus millal küsitleda, rakendati jagamisalgoritmi. Algoritmi jaotab asutused määratud numbrilise tunnuse järgi nii, et selle tunnuse kogusumma mõlemas pooles oleks võimalikult sarnane. Vaadeldi eraldi kahte tunnust: isikute arvu ja objektide arvu asutuses, kus objektiks on isik erinevatel ametitel (andmetabeli rida). Selgus, et tulemused ei sõltu jagamisel kasutatavatest tunnustest, mistõttu tulevikus piisab algoritmi realiseerimisel objektide arvu põhise jagamisviisi kasutamisest. Seda tuleks eelistada ka selle lihtsuse tõttu, kuna objektide arv on sama, mis ridade arv.

Teiseks väljakutseks magistritöös vaadeldud meetodite rakendamisel oli aastatega muutuv andmestik. Uute asutuste jaoks tuli määrata, millisel aastal nendelt aruannet küsida. Küsitlemis-aasta valiti lähtuvalt sellest, kui palju asutusi ja milliste töötajate või objektide arvudega, oli vaadeldaval aastal lõpetanud tegevuse või ei töötanud.

Käesoleva töö raames uuriti kolme parameetri hindamist: tervishoiutöötajate arv, tegelikult täidetud ametikohtade arv ja keskmine vanus.

Tervishoiutöötajate arvu hindamisel andsid täpseimaid tulemusi veereva disainiga 2 saadud hinnangud, kus vaadeldava aasta kaal oli 60%. Tegelikult täidetud ametikohtade hindamisel andsid täpseimaid tulemusi veereva disainiga 1 saadud hinnangud, kus vaadeldava ja sellele eelneva aasta kaalumiseks kasutati ainult tervishoiutöötajate arvust tulenevaid kaalusid.

Keskmise vanuse hindamiseks kasutatud kolme meetodi hinnangud olid kõik sarnase täpsusega. Kuna keskmine vanus sõltub tervishoiutöötajate arvu hinnangust, mille jaoks oli parim veerev disain 2, siis ka keskmise vanuse jaoks on mõttekas neist valida veerev disain 2.

Kokkuvõttes võib järeldata, et parimaks antud töös uuritud meetodiks on veerev disain 2. Parema tulemus on tingitud sellest, et suuri asutusi (haiglaid) küsitletakse igal aastal.

TAI otsib aktiivselt võimalusi nii asutuste kui ka andmekogujate koormuse vähendamiseks. Käesolevas magistritöös uuritud meetodid tõestasid, et asutusi on võimalik küsitleda osaliselt, säilitades samal ajal vajaliku statistilise informatsiooni ja selle kvaliteedi. Antud töö raames ei uuritud, milliseid tulemusi annavad vaadeldud meetodid erinevate osakogumite lõikes.

Meetodite käitumist on võimalik veelgi parandada, uurides teisi viise asutuste jagamiseks ja kattuva osa väljavalimiseks. Ei ole välistatud teiste statistiliste, näiteks aegridadel põhinevate meetodite kasutamine.

## Kasutatud kirjandus

- [1] Anderson, E. (2019). Imikute rinnapiimaga toitmise statistika kvaliteediraport. Tallinn: Tervise Arengu Instituut.
- [2] Tervishoiuteenuste korraldamise seadus (09 05 2001). *Riigi Teataja I*. Kasutatud 10.03.2020, <https://www.riigiteataja.ee/akt/109122016012?leiaKehtiv>
- [3] Tervisestatistika ja terviseuuringute andmebaasi lehekülg. (i.a.). Kasutatud 20.05.2020, <https://statistika.tai.ee/>
- [4] Kish, L. (1990). Rolling Samples and Censuses. *Survey Methodology* , 16, 63-79.
- [5] Särndal, C.-E., Swensson, B., Wretman, J. (2003). Model Assisted Survey Sampling, New York:Springer.
- [6] Tervise Arengu Instituut: tervishoiutöötajate ametite kodeerimine. (i.a.). Kasutatud 17.02.2020, [https://www.tai.ee/images/PDF/Klassifikaatorid/Tervishoiut%C3%B6%C3%B6tajate\\_ametite\\_kodeerimine.pdf](https://www.tai.ee/images/PDF/Klassifikaatorid/Tervishoiut%C3%B6%C3%B6tajate_ametite_kodeerimine.pdf)
- [7] Eesti Statistika: Eesti haldus- ja asustusjaotuse klassifikaator. (2019). Kasutatud 17.02.2020, [http://metaweb.stat.ee/view\\_xml.htm?id=4601354&siteLanguage=ee](http://metaweb.stat.ee/view_xml.htm?id=4601354&siteLanguage=ee)
- [8] Tervise Arengu Instituut: tervishoiuteenuse osutajate liigitus. (i.a.). Kasutatud 17.02.2020, [https://www.tai.ee/images/PDF/Klassifikaatorid/Tervishoiuteenuse\\_osutajate\\_liigitus.pdf](https://www.tai.ee/images/PDF/Klassifikaatorid/Tervishoiuteenuse_osutajate_liigitus.pdf)
- [9] Tervise Arengu Instituut: ICHA-HP Tervishoiuteenuse pakkujate liigitus. (i.a.). Kasutatud 17.02.2020, [https://www.tai.ee/images/PDF/Klassifikaatorid/SHA2011\\_ICHA-HP.pdf](https://www.tai.ee/images/PDF/Klassifikaatorid/SHA2011_ICHA-HP.pdf)
- [10] Kish, L. (1998). Space/Time Variations and Rolling Samples. *Journal of Official Statistics*, 14(1), 31-46.
- [11] Alexander, C. H. (2002). Still Rolling: Leslie Kish's "Rolling samples" and the American Community Survey. *Survey Methodology* , 28(1), 35-41.

## Lisa 1. R kood asutuste pooleks jagamiseks

Kõigepealt loeme sisse vajalikud paketid. Käivitame kommentaarides olevad koodid, kui antud paketid pole varem alla laetud.

```
#install.packages("haven")
#install.packages("dplyr")
#install.packages("reshape2")
library(haven)
library(dplyr)
library(reshape2)
```

Kasutame andmestikku `asutused`, kus igale asutusele on määratud selle kiht ja tunnus, mille järgi soovitakse asutusi jagada. Käesolevas töös korraldame seda algoritmi kaks korda: esimesel korral oli tunnuseks tegevishoiutöötajate arv asutuses ja teisel korral objektide (isik-amet) arv asutuses.

Jätame andmestikku vaid vajalikud tunnused ja loome tunnuse `pool`, mis on esialgu tühi. Jagamisel kasutatava tunnuse nimeks on `tunnus`. Lisaks defineerime eraldi vektori `kihid`, kus on kihtide väärtuste loetelu.

```
asutused<-asutused %>% select(id_kood, kiht, tunnus)%>%
  arrange(kiht,tunnus) %>% mutate(pool=0)

kihid<-sort(unique(asutused$kiht))
kihid

## [1] 11 12 13 14 15 16 17 20 25 30 40 50 51 53 100
```

Koostatakse tsükkel üle kõikide kihtide.

```
for(i in kihid){
  NN=length(asutused$id_kood[asutused$kiht==i])
  N=NN
  if(NN==2){
    asutused[asutused$kiht==i,$pool[1]]=1
    asutused[asutused$kiht==i,$pool[2]]=2
  }else{
    viimane<-asutused[asutused$kiht==i,$tunnus[N]
    eelviimased<-sum(asutused[asutused$kiht==i,$tunnus[1:(N-1)])
    if(eelviimased < viimane){
      asutused[asutused$kiht==i,$pool[N]<-1
      asutused[asutused$kiht==i,$pool[1:(N-1)]<-rep(2,length(1:(N-1)))
    } else if(NN==4){
      asutused[asutused$kiht==i,$pool[c(1,NN)]<-rep(1,2)
      asutused[asutused$kiht==i,$pool[c(2,3)]<-rep(2,2)
    }else{
      asutused[asutused$kiht==i,$pool[1]<-1*(NN%%2==0)+2*(NN%%2!=0)
      k=2
    }
  }
}
```



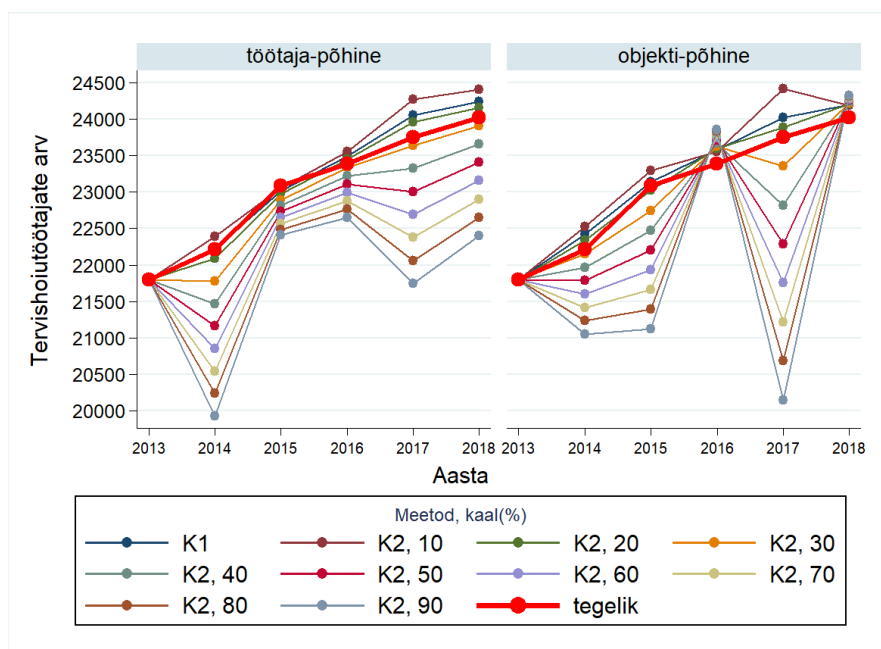
```

while(sum((asutused[asutused$kiht==i,]$pool)==0)>0){
  j=k
  asutused[asutused$kiht==i,]$pool[N]<-1
  esimenaryhm=sum(asutused[asutused$kiht==i,]$tunnus[
    asutused[asutused$kiht==i,]$pool==1&
    (asutused[asutused$kiht==i,]$pool!=0)])
  teineryhm=sum(asutused[asutused$kiht==i,]$tunnus[
    asutused[asutused$kiht==i,]$pool==2&
    (asutused[asutused$kiht==i,]$pool!=0)])
  eelviimased<-sum(asutused[asutused$kiht==i,]$tunnus[j:(N-1)])
  while((eelviimased+teineryhm)>esimenaryhm){
    j=j+1
    eelviimased<-sum(asutused[asutused$kiht==i,]$tunnus[j:(N-1)])
    if(j>=N-1) next
  }
  asutused[asutused$kiht==i,]$pool[c(k,j:(N-1))]<-
    rep(2,length(c(k,j:(N-1))))
  if(k==j) break
  asutused[asutused$kiht==i,]$pool[k+1]<-1
  N=j-1
  k=k+2
  if(N==k) {
    asutused[asutused$kiht==i,]$pool[k]<-2
    break
  }
}
}
}
}
}

```

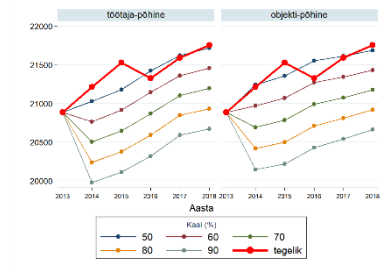
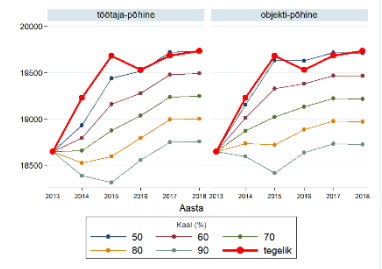
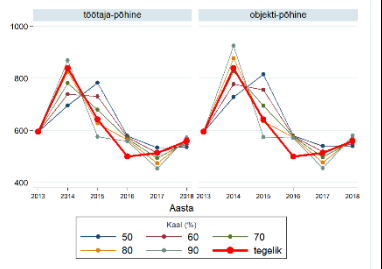
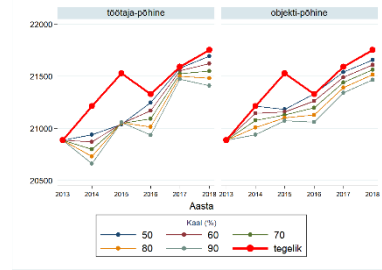
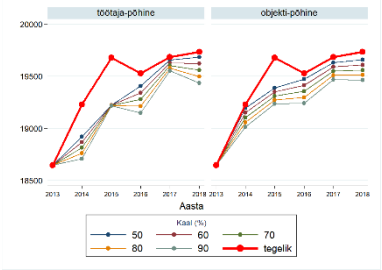
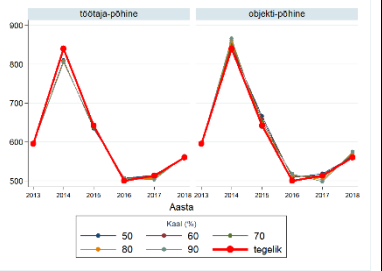
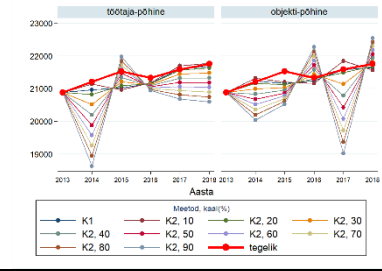
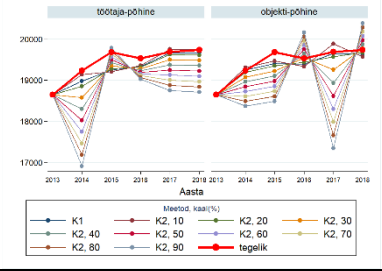
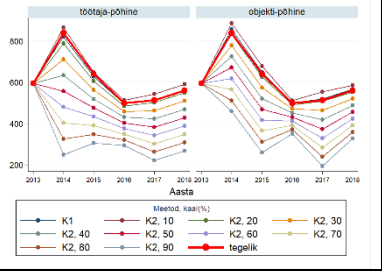
Selle jagamisalgoritmi tulemuseks on tunnus pool väärtustega 1 ja 2 sõltuvalt sellest, millisesse poolde asutus sattus.

## Lisa 2. Hinnangud erinevate meetodite ja tunnuste korral



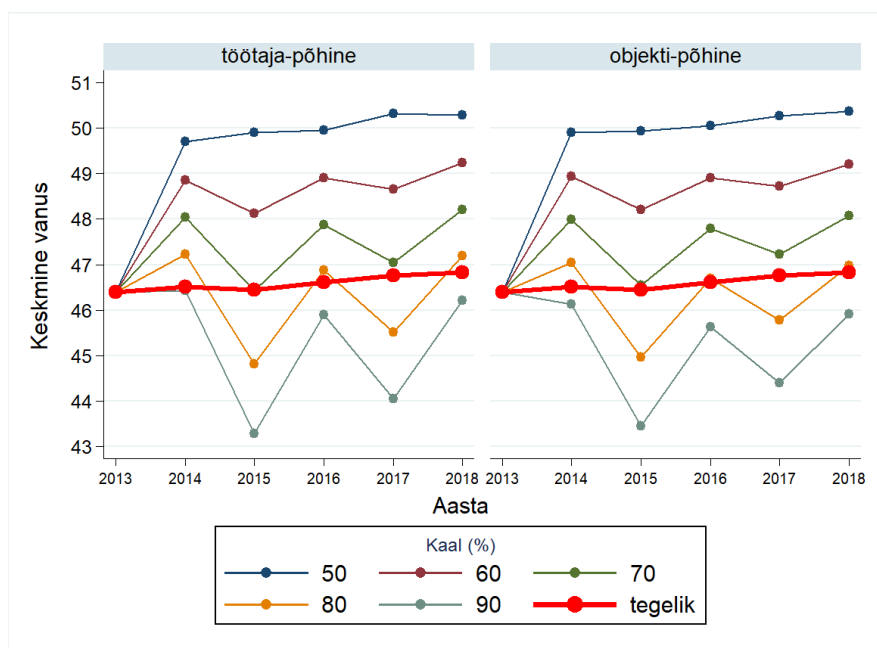
Joonis L1. Tervishoiutöötajate hinnangud kahe hetke disaini 1 ja 2 korral

Tabel L1. Eri liiki koormustest tulenevad ametikohtade hinnangud

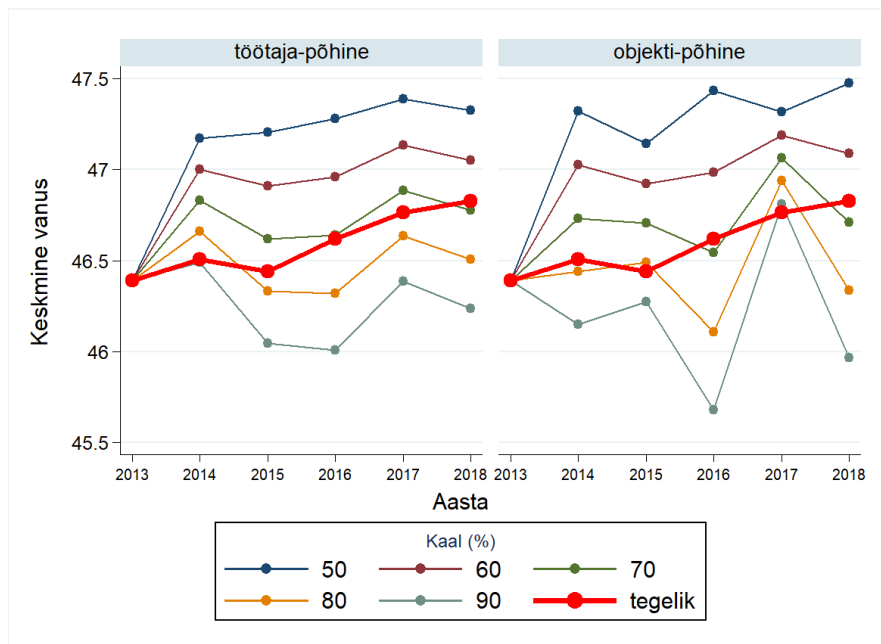
Täidetud ametikohtade arvu hinnangud lepingulise koormuse järgi	Tegelikult täidetud ametikohtade arvu hinnang	Ületundidest tulenev täiendav täidetud ametikohtade arvu hinnang
1. Veerev disain 1		
		
1. Veerev disain 2		
		
2. Kahe hetke disaini 1 ja 2 hinnangud		
		

Tabel L2. Eri liiki koormustest tulenevate ametikohtade hinnangute lähendusvead *RMSE* jagamisviisi ja meetodi lõikes

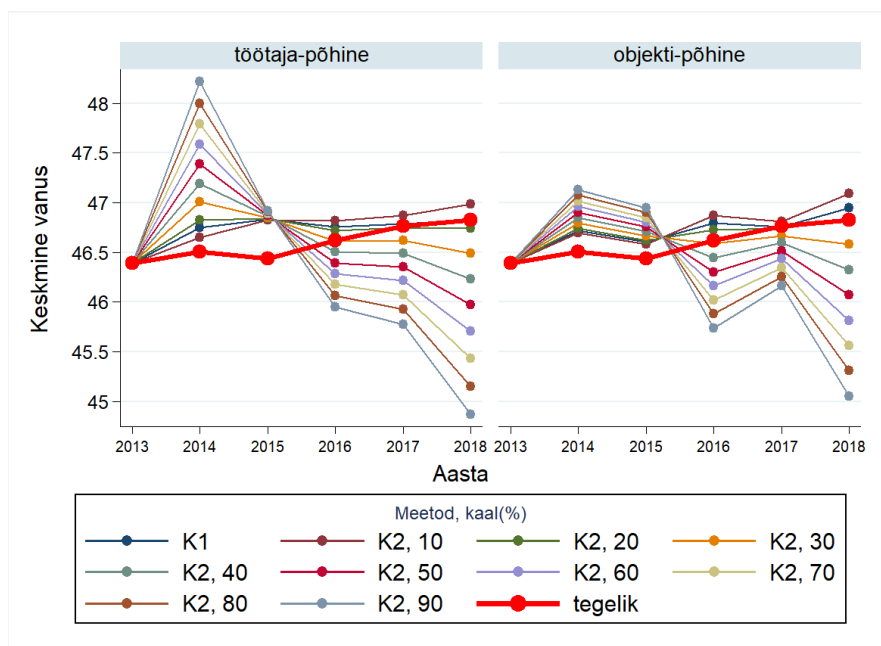
Parameeter		Täidetud ametikohtade arvu hinnangud lepingulise koormuse järgi		Tegelikult täidetud ametikohtade arvu hinnang		Ületundidest tulenev täiendav täidetud ametikohtade arvu hinnang	
Jagamise viis		Töötaja-p.	Objekti-p.	Töötaja-p.	Objekti-p.	Töötaja-p.	Objekti-p.
Meetod	Kaalud						
Veerev disain 1	50	138,1	103,6	117,6	55,5	81,4	82,5
	60	353,6	264,5	332,0	241,3	55,7	53,9
	70	618,3	537,1	560,8	477,8	37,7	31,5
	80	8823,0	809,7	789,5	714,3	25,0	32,8
	90	1147,7	1082,3	1018,2	950,9	45,1	60,4
Veerev disain 2	50	182,5	99,2	193,2	102,5	9,3	8,9
	60	230,5	149,4	234,5	148,5	9,1	8,6
	70	278,5	201,8	275,8	194,5	9,7	10,3
	80	326,5	254,2	317,2	240,5	10,2	12,2
	90	374,6	306,6	358,5	286,5	10,8	15,3
Kahe hetke disain 1		203,0	118,8	196,7	115,5	7,9	8,0
Kahe hetke disain 2	10	183,6	205,2	163,2	169,9	21,4	32,5
	20	228,5	128,7	233,0	127,6	24,1	8,1
	30	326,9	267,2	333,3	218,8	69,5	48,6
	40	425,2	439,2	433,5	371,4	115,0	89,2
	50	523,5	611,1	533,8	524,0	160,5	129,8
	60	648,6	783,0	634,1	676,6	206,0	170,4
	70	797,5	954,9	734,4	829,2	251,4	210,9
	80	946,5	1126,9	848,5	981,8	296,9	251,5
	90	1095,5	1298,8	977,7	1134,5	342,4	292,1



Joonis L2. Hinnang keskmisele vanusele veereva disaini 1 korral



Joonis L3. Hinnang keskmisele vanusele veereva disaini 2 korral



Joonis L4. Hinnang keskmisele vanusele kahe hetke disaini 1 ja 2 korral

## **Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks**

Mina, Diana Sokurova,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Tervishoiutöötajate aruandluskoormuse vähendamine veereva ja kahe hetke valimidisaini abil ja kahe hetke valimidisaini abil“,

mille juhendajad on Imbi Traat ja Natalja Eigo,

reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

*Diana Sokurova*  
**25.05.2020**